

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 14:29:45 ; Search time 57.98 Seconds  
(without alignments)  
8922.114 Million cell updates/sec

Title: US-09-856-327-1  
Perfect score: 2106  
Sequence: 1 atcagcatgtctctctca.....aaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents,NA:\*  
1: /cgn2\_6/pdata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/pdata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/pdata/1/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/pdata/1/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/pdata/1/lna/PCUS\_COMB.seq:\*  
6: /cgn2\_6/pdata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	212.4	10.1	1869	3	US-09-305-381-1
2	210.2	10.0	1869	1	US-08-734-925-1
3	152.6	7.2	1701	4	US-09-023-731-2
4	88.8	4.2	1733	3	US-09-073-569-1
5	87	4.1	2239	4	US-09-196-390-1
6	86.6	4.1	2671	6	5168051-9
7	85	4.0	1582	3	US-08-545-196B-10
8	84.4	4.0	1582	3	US-08-545-196B-12
9	84.4	4.0	1641	1	US-08-300-903A-8
10	84.4	4.0	3410	4	US-09-020-956-110
11	84.4	4.0	3410	4	US-09-030-607-110
12	84.4	4.0	3410	4	US-09-439-313-110
13	83.6	4.0	1474	4	US-08-821-994-64
14	83.4	4.0	1602	1	US-08-530-950-3
15	83.4	4.0	1602	3	US-08-888-429A-3
16	83.4	4.0	1602	4	US-09-149-879-3
17	83.2	4.0	2852	3	US-09-027-137-2
18	83	3.9	1098	3	US-09-248-335-35
19	83	3.9	1134	3	US-09-248-335-29
20	83	3.9	1193	4	US-09-372-422A-23
21	82.6	3.9	1813	4	US-09-071-224-3
22	82.4	3.9	144	1	US-08-702-344-26
23	82.4	3.9	1117	4	US-09-247-373B-33
24	82.4	3.9	1307	2	US-08-960-022-17
25	82.4	3.9	2082	2	US-08-785-310A-2
26	82.2	3.9	1736	3	US-09-182-816-22
27	82.2	3.9	1736	3	US-09-182-816-24

c

ALIGNMENTS

RESULT 1

US-09-305-381-1  
; Sequence 1, Application US/09305381  
; Patent No. 6145865  
; GENERAL INFORMATION:  
; APPLICANT: Christensen, Soren  
; APPLICANT: Lassen, Soren Flensted  
; APPLICANT: Schneider, Falle  
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides  
; TITLE OF INVENTION: Having Pyranose Oxidase Activity  
; FILE REFERENCE: 5571.200-US  
; CURRENT APPLICATION NUMBER: US/09/305,381  
; CURRENT FILING DATE: 1999-05-05  
; EARLIER APPLICATION NUMBER: 60/088,724  
; EARLIER FILING DATE: 1998-06-10  
; EARLIER APPLICATION NUMBER: PA 1998 00774  
; EARLIER FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1869  
; TYPE: DNA  
; ORGANISM: Trametes hirsuta  
US-09-305-381-1

Query Match 10.1%; Score 212.4; DB 3; Length 1869;  
Best Local Similarity 51.5%; Pred. No. 6.5e-41;  
Matches 841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;

QY	269	ggctacacaaagaagaatgaatcgaggtctccagaagaagattaccgctctcgcaatgta	328
Db	259	ggctacacaaagaagaacaccgctcgagtagcagaagaacatcgacaaatcgtaattgt	318
QY	329	atcaaggagcccttacaacaaagtctctgtctcgtcagaaccagacgctgcacactt	388
Db	319	atcaaggagccacttatgccgctctcgtgcccgtcgaacacgatggtcgttgacacgcta	378
QY	389	gatcccgagccttgagcgcgcgcccttggagttcagccatctcgaacgtaaaaaatcct	448
Db	379	agccccggtcactggcaagtctcgacg-----ttcttcgcgaacggcgcaatcca	432
QY	449	cacacagcgggaattcagaaacttgctgcggagcgttaacgcgttgagtcggcgcatg	508
Db	433	gagcagacccgctgcgaaccttagtggcagcgggtaccgcgctcgtcggcgcatg	492
QY	509	agtaccactggagctgctccacgcacgattcattccaccatggaagtctcccgggc	568
Db	493	tctacgcactggagctgcgcgcgcgctcttcgagaagctgcag-----	537





```

, FILING DATE: 13-FEB-1998
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: JP 29275/1997
, FILING DATE: 13-FEB-1997
, ATTORNEY/AGENT INFORMATION:
, NAME: KENNETH H. SONNENFELD
, REGISTRATION NUMBER: 33,285
, REFERENCE/DOCKET NUMBER: 3479-4000
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (212) 758-4800
, TELEFAX: (212) 751-6849
, TELEX: 421782
, INFORMATION FOR SEQ ID NO: 2:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1701
, TYPE: nucleic acid
, TOPOLOGY: linear
, STRANDEDNESS: Single
, MOLECULE TYPE: cDNA to RNA
US-09-023-731-2

```

Query Match 7.2%; Score 152.6; DB 4; Length 1701;  
Best Local Similarity 53.5%; Pred. No. 7.1e-27;  
Matches 394; Conservative 0; Mismatches 304; Indels 39; Gaps 2;

QY	1106	ctgggagagbtacatcacggagcagcggatggcattttgcccagatagctcttgagcaggaa	1165
Db	964	CTTGGTCGTCTACTCAGCGAAACAGTGCCTATGACTTTTGTGCAGATCGTTCTCAAGAGGGC	1023
QY	1166	ttcgctgcagcgtgcgagcagctcttatggactgccatggtggaagaagccgttgct	1225
Db	1024	ATAGTCGATGCCATCGCTACTGACCTT-----CGCTTCGCTGGAGGTTGAG	1071
QY	1226	caacatattgccaagaaccgcagagatcactgcccattccgttccgcgcatccggaaacc	1285
Db	1072	GCGCACAAGAAGACACCCCGATCAGCTGTGTGCCATTCCATTCCACGAGCCTGAACCT	1131
QY	1286	caggtaaacaccccatttacagaagaacacccctggcacgcagattccaccgatgct	1345
Db	1132	CAGTGATGATTCGGTACACGTCGGGACTTCCCTTGGCATGTTTCAGGTGCTACGCGATGCA	1191
QY	1346	tttctcaggtgcogtgcgtcagtgaggtagctctcgtgcatcgtcgaccctgcgctgg	1405
Db	1192	TTCTCATATGGTGATGTGGACCCAAAGCCGACCCGGCTGTGTGCTGATCTGAGGTTT	1251
QY	1406	tttggcgcaaccgaccctgaagcaaacacatttggtttc-----	1447
Db	1252	TTCCGGCAATCAGATATTGTCTGAAGAAATCGATGACTTTTCGGTCCGAACCTTAGCTA	1311
QY	1448	-----cagaacgatgttcaagcaggttacagtagtcgcgcagcgcagcttcagatat	1498
Db	1312	CGCGAGTGGGAAGCGGGTGTTACAGACACTTATGSAATGCCACAGCCGACATCCATGTC	1371
QY	1499	cgaccagcaactcgtcaaacgtgagagcaaggaaaaatgatgcccgatatgtgcgaagt	1558
Db	1372	AGCGGACCAACGCCGATGGAGACCGTGACCAGAGGATGATGAATGATATGACCACAGTC	1431
QY	1559	gcagcaacttgggaggttattgcccagctcccccccgagttatggatccaggcctt	1618
Db	1432	CGCAACATGCTGGGTGGTACCTTCTTGGCTCCATACCTCAATTTATGGCACCTGGTCTC	1491
QY	1619	gcaacttcatttgcggggactactcgaattggcttcgacaaggcaactcagtggtgat	1678
Db	1492	GTACTGCATCAGTCGGGAACACTACTCGGATCGGGACAGATGATCAAACTTCTGTTGCTGAT	1551
QY	1679	aaccaactcgtggtctgggactttgccaatcttatgttgaggcaaatgcacacatcagg	1738
Db	1552	CCGACATCAAGGTTCACTTCAACAAATCTGTGGGTGCGGGGAATGGTGCAATCCA	1611
QY	1739	acgggtctcggcgaaacccgacacttacgtcgaatgtgccacgctatcaagagcgcgagg	1798
Db	1612	GATGCGACTGCTGCAACCCGACTCGTACAGAGCTCGCGTATGGCTTCAAGGGTGCTGAG	1671

```

Qy 1799 agcatcatcaatcaact 1815
      | | | | |
Db 1672 GCTGTAGTCAATTACCT 1688

RESULT 4
US-09-073-569-1
: Sequence 1, Application US/09073569
: Patent No. 6084088
: GENERAL INFORMATION:
:   APPLICANT: Sheppard, Paul O.
:   APPLICANT: Grossmann, Angelika
:   TITLE OF INVENTION: NOVEL TUMOR ANTIGENS
:   NUMBER OF SEQUENCES: 15
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: ZymoGenetics, Inc.
:   STREET: 1201 Eastlake Avenue East
:   CITY: Seattle
:   STATE: WA
:   COUNTRY: USA
:   ZIP: 98102
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Diskette
:   COMPUTER: IBM Compatible
:   OPERATING SYSTEM: DOS
:   SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/073,569
:   FILING DATE:
:   CLASSIFICATION:
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER:
:   FILING DATE:
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Sawislak, Deborah A
:     REGISTRATION NUMBER: 37,438
:     REFERENCE/DOCKET NUMBER: 97-14
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 206-442-6672
:   TELEFAX: 206-442-6678
:   TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1733 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   FEATURE:
:   NAME/KEY: Coding Sequence
:   LOCATION: 34...1344
:   OTHER INFORMATION:
US-09-073-569-1

```

Query Match 4.2%; Score 88.8; DB 3; Length 1733;  
Best Local Similarity 79.5%; Pred. No. 7.1e-12;  
Matches 105; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy	1775	tattactgaacctgtactttgtctgtaataagttatggcactatgatctcatgtttaaaaaaa	2034
Db	1583	TAATACAGACCCTGTCTTTCCTCCCAAAAAAATAAAAAAAAAAAAAAAAAAAAAAA	1642
Qy	2035	aaa	2094
Db	1643	AA	1702
Qy	2095	aaaaaaaaaaaaaaa	2106
Db	1703	AAAAAAAAAAAAAAA	1714

## 5 RESULT



US-09-196-390-1  
; Sequence 1, Application US/09196390  
; Patent No. 6307125  
; GENERAL INFORMATION:  
; APPLICANT: Block, Martina  
; APPLICANT: Lorz, Horst  
; APPLICANT: Lutticke, Stephanie  
; APPLICANT: Walter, Lennart  
; APPLICANT: Froberg, Claus  
; APPLICANT: Kossmann, Jens  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES  
; TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH  
; TITLE OF INVENTION: SYNTHESIS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/196,390  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 196 21 588.9  
; FILING DATE: 29-MAY-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 196 36 917.7  
; FILING DATE: 11-SEP-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP97/02793  
; FILING DATE: 28-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: AGREVO-9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2239 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Triticum aestivum L.  
; STRAIN: cv. Florida  
; HAPLOTYPE: ca. 21 d Caryopses  
; IMMEDIATE SOURCE:  
; LIBRARY: cDNA library in pBluescript sk (-)  
; CLONE: TASSS  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3..2017  
US-09-196-390-1

Query Match 4.1%; Score 87; DB 4; Length 2239;  
Best Local Similarity 67.2%; Pred. No. 2.1e-11;  
Matches 123; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 1924 tctgtccctaccatgtgatgtacatagcgttgaaagattttgtgtattactga 1983

Db 2054 TGTGTCTGTGCTTAGCTGACAAATATTTGACCTGTTGGAGAAATTTTATCTTTGCTGCT 2113  
Qy 1984 acctgtacttctcgaatgattgacacatgatcatgttttaaaaaa 2043  
Db 2114 GTTTTTTTTAATCAAAAGAGGGGTTTCCTCCGATTTTCATTA 2173  
Qy 2044 aa 2103  
Db 2174 AA 2233  
Qy 2104 aaa 2106  
Db 2234 AAA 2236

## RESULT 6

5168051-9  
; Patent No. 5168051  
; APPLICANT: DERYNCK, RIK M.A.; GOEDEL, DAVID V.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES  
; NUMBER OF SEQUENCES: 21  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/389,929  
; FILING DATE: 04-AUG-1989  
; SEQ ID NO: 9  
; LENGTH: 2671  
5168051-9

Query Match 4.1%; Score 86.6; DB 6; Length 2671;  
Best Local Similarity 83.8%; Pred. No. 2.7e-11;  
Matches 98; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1990 acttctgtcgaatgattgacacatgatcatgttttaaaaaa 2049  
Db 2544 attgtcttaacataaagccttattcgaagtgttaaaaaa 2603  
Qy 2050 aa 2106  
Db 2604 aa 2660

## RESULT 7

US-08-545-196B-10  
; Sequence 10, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUNNICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,196B  
; FILING DATE: 19-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FARACI, C. J.  
; REGISTRATION NUMBER: 32,350  
; REFERENCE/DOCKET NUMBER: 2121-110P

Qy	1981	tgaacctgtacttctgctgaatgtagttatgtagcactatgattcatctgtttataaaaaaaaaa	2040
Db	1508	TGACATGTGACCTTGATTAACACAAATAAACAATCAATCTATTTTCAATAAAAAAAAAA	1567
Qy	2041	aaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa	2100
Db	1568	AA	1627
Qy	2101	aaaaaa 2106	

```
Db 1628 AAAAA 1633
|||||
RESULT 10
US-09-020-956-110
; Sequence 110, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-09-020-956-110

Query Match 4.0%; Score 84.4; DB 4; Length 3410;
Best Local Similarity 80.8%; Pred. No. 9.9e-11;
Matches 97; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY 1987 tgtactttgtcgaatagtcagtcactatgattctgttataaaataaaataaaataaa 2046
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3287 TGTATTGTGTCACAAATTAAGGCTTTCTATATGTTTAAATAAAATAAAATAAA 3346
QY 2047 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAAATAAA 3406

RESULT 11
US-09-030-607-110
; Sequence 110, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
```

```
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-09-030-607-110

Query Match 4.0%; Score 84.4; DB 4; Length 3410;
Best Local Similarity 80.8%; Pred. No. 9.9e-11;
Matches 97; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY 1987 tgtactttgtcgaatagtcagtcactatgattctgttataaaataaaataaa 2046
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3287 TGTATTGTGTCACAAATTAAGGCTTTCTATATGTTTAAATAAAATAAAATAAA 3346
QY 2047 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAA 3406

RESULT 12
US-09-439-313-110
; Sequence 110, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
;
US-09-439-313-110
```

```
Query Match 4.0%; Score 84.4; DB 4; Length 3410;
Best Local Similarity 80.8%; Pred. No. 9.9e-11;
Matches 97; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

Qy 1987 tgtacttgctgaatgttatggtgacatgatgattcattgtttataaaaaa 2046
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3287 tggttatggtgacaaaataaaggcttcttattgtttataaaaaa 3346

Qy 2047 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3347 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3406

RESULT 13
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64

Query Match 4.0%; Score 83.6; DB 4; Length 1474;
Best Local Similarity 75.4%; Pred. No. 1.1e-10;
Matches 104; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1969 ttgtgtattactgaacctgtactttgtctgaatagttatggcactatgattcgttta 2028
||| |||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1333 ttatgtattaagagagataataaaatgatattcttcttaaaaaa 1392

Qy 2029 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2088
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1393 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1452

Qy 2089 aaaaaaaaaaaaaaaaaa 2106
||||| ||||||| |||
Db 1453 aaaaaaaaaaaaaaaaaa 1470

RESULT 14
US-08-530-950-3
; Sequence 3, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
```

```
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-530-950-3

Query Match 4.0%; Score 83.4; DB 1; Length 1602;
Best Local Similarity 93.5%; Pred. No. 1.3e-10;
Matches 87; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2014 tatgtattcatgtttataaaaaa 2073
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1509 TATGAGTCCTCAAAAAA 1601

Qy 2074 aaaaaaaaaaaaaaaaaa 2106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1569 AAAAAA 1601

RESULT 15
US-08-888-429A-3
; Sequence 3, Application US/08888429A
; Patent No. 6136596
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
; TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE KINASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429A
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
```

```

/ FILING DATE: 19-MAY-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fasse, Peter J.
/ REGISTRATION NUMBER: 32,983
/ REFERENCE/DOCKET NUMBER: 07917/053001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617/542-5070
/ TELEFAX: 617/542-8906
/ TELEX: 299354
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1602 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: Coding Sequence
/ LOCATION: 244...1245
/ US-08-888-429A-3

```

Query Match 4.0%; Score 83.4; DB 3; Length 1602;  
Best Local Similarity 93.5%; Pred. No. 1.3e-10;  
Matches 87; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

[illegible]

Qy	2074	aaaaa	aaaaaaaaaaaaaaaaaaaaaaaa	2106
D6 <td>1569</td> <td>AAAAA</td> <td>AAAAAAAAAAAAAAAAAAAAAA</td> <td>1601</td>	1569	AAAAA	AAAAAAAAAAAAAAAAAAAAAA	1601

Search completed: September 3, 2002, 15:51:01  
Job time: 4876 sec

***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 14:36:36 ; Search time 242.59 Seconds  
(without alignments)  
14905.074 Million cell updates/sec

Title: US-09-856-327-1  
Perfect score: 2106  
Sequence: 1 atcagcatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2106	100.0	2106	22	AAF99980 Nucleotide sequenc
2	212.4	10.1	1869	22	AAH87519 Trametes hirsuta p
3	212.4	10.1	1935	22	AAH87518 Trametes hirsuta p
4	208.6	9.9	1869	17	AAH34420 Pyranose oxidase e
5	175.6	8.3	1902	21	AAZ46411 Pleurotus cornucop
6	152.6	7.2	1701	20	AAV83626 Nucleic acid encod
7	129	6.1	1946	21	AAH71487 T. matsutake pyran
8	129	6.1	1946	21	AAH71487 Trichoderma derive
9	94.4	4.5	4639	22	AAH88688 Human DNA encoding

10	94.4	4.5	4640	21	AAH78483 Human PRO708 (UNQ3
11	94.4	4.5	4650	20	AAH33987 Human PRO708 nucle
12	91.8	4.4	2440	22	AAH34932 Human colon cancer
13	91.6	4.3	2323	19	AAV9524 Human secreted pro
14	91.4	4.3	1480	22	AAH07771 Human secreted pro
15	91.2	4.3	2710	21	AAH54132 Breast cancer prot
16	90.8	4.3	1327	24	AAH24775 Glycine max ankyri
17	90.6	4.3	1392	22	AAH72748 Human prostate can
18	89.6	4.3	374	22	AAH91055 Human polynucleoti
19	89.6	4.3	1046	24	AAH63134 Cell death protect
20	89.6	4.3	2260	22	AAH97914 Human neuroblastom
21	89.6	4.3	2260	22	AAH98068 Human neuroblastom
22	89.4	4.2	1768	22	AAH08093 Human breast cancer
23	89.2	4.2	1735	22	AAH08093 Human breast cancer
24	88.8	4.2	1733	20	AAH81394 Human tumour antig
25	88.4	4.2	400	22	AAH84784 Human polynucleoti
26	88.4	4.2	1091	22	AAH89723 Maize ZmGnsN1-1 g1
27	88.2	4.2	421	22	AAH91775 Human polynucleoti
28	88	4.2	422	22	AAH87735 Human polynucleoti
29	88	4.2	425	22	AAH60450 Human cancer agent
30	88	4.2	567	22	AAH33915 Human colon cancer
31	88	4.2	1248	21	AAH26436 Human secreted pro
32	87.8	4.2	310	22	AAH71505 Human cervical can
33	87.8	4.2	823	22	AAH07664 Human ovarian and
34	87.8	4.2	823	22	AAH07664 Human reproductive
35	87.8	4.2	3312	8	AAH71034 pmx34 cDNA insert.
36	87.6	4.2	393	22	AAH11189 Human breast cancer
37	87.6	4.2	10039	24	ABL34042 Human immune syste
38	87.4	4.2	545	22	AAH70126 Human cervical can
39	87.4	4.2	546	22	AAH71167 Human cervical can
40	87.4	4.2	1199	20	AAH08695 Novel nucleotide s
41	87.2	4.1	765	21	AAH59300 Human secreted pro
42	87.2	4.1	1254	22	AAH25886 Human cDNA encodin
43	87.2	4.1	2434	21	AAH26402 Human secreted pro
44	87	4.1	379	22	AAH20476 Human breast cancer
45	87	4.1	392	22	AAH88688 Human polynucleoti

ALIGNMENTS

RESULT 1  
AAF99980 standard; cDNA; 2106 BP.

XX AAF99980;

XX 20-JUL-2001 (first entry)

XX Nucleotide sequence encoding Lyophyllum shimeji antibacterial protein.  
XX Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;  
XX Pyricularia oryzae; Rhizoctonia solani; rice pathogen; ss.

XX Lyophyllum shimeji.

XX Key Location/Qualifiers

XX CDS 8..1864

XX FT /\*tag- a

XX FT /product= "antibacterial protein"

XX WO200121657-A1.

XX 29-MAR-2001.

XX 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NIBS ) JAPAN TOBACCO INC.

XX (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTRY & FI.

XX Takakura Y, Kuwata S, Inoue Y;





Db 1741 gggcttcggcgagaccgcacactacgtcgtgatgtgccacgtctatcaagagcgcgaggag 1800  
 QY 1801 catcataactacactcaagggtggactgacgggaaataatcacagcgagcatcgcaacct 1860  
 Db 1801 catcataactacactcaagggtggactgacgggaaataatcacagcgagcatcgcaacct 1860  
 QY 1861 ttgaggaagagcaacagcagtgtaaacaaacgcgtcaagtggtcactctcaagttaagt 1920  
 Db 1861 ttgaggaagagcaacagcagtgtaaacaaacgcgtcaagtggtcactctcaagttaagt 1920  
 QY 1921 cattctggctccctaccattgtatgtgtacgatagcggtgaaagatttggattac 1980  
 Db 1921 cattctggctccctaccattgtatgtgtacgatagcggtgaaagatttggattac 1980  
 QY 1981 tgaactgtactttgtctgaatgtatgtgacactatgattcattgtttaaaaaaaaaa 2040  
 Db 1981 tgaactgtactttgtctgaatgtatgtgacactatgattcattgtttaaaaaaaaaa 2040  
 QY 2041 aa 2100  
 Db 2041 aa 2100  
 QY 2101 aaaaaa 2106  
 Db 2101 aaaaaa 2106

RESULT 2  
 AAC87519  
 ID AAC87519 standard; DNA; 1869 BP.  
 XX AC AAC87519;  
 XX DT 13-MAR-2001 (first entry)  
 XX DE Trametes hirsuta pyranose oxidase cDNA, SEQ ID NO:1 (version 2).  
 XX KW Pyranose oxidase; expression construct; recombinant production;  
 KW monosaccharide oxidation; 2-keto derivative;  
 KW hydrogen peroxide production; ss.  
 XX OS Trametes hirsuta.  
 XX PN US6146865-A.  
 XX PD 14-NOV-2000.  
 XX PF 05-MAY-1999; 99US-0305381.  
 XX PR 08-JUN-1998; 98DK-0000774.  
 XX PR 10-JUN-1998; 98US-0088724.  
 XX PA (NOVO ) NOVO NORDISK AS.  
 XX PI Schneider P, Christensen S, Lassen SF;  
 XX WPI: 2001-049055/06.  
 XX P-PSDB: AAB48832.  
 XX DR Novel nucleic acid molecule encoding polypeptide having pyranose  
 PT oxidase activity used to design oligonucleotide probes to identify and  
 PT clone DNA encoding the polypeptide from different genera or species -  
 XX Claim 2; Column 25-28; 20pp; English.  
 XX PS  
 CC The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which  
 CC encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also  
 CC relates to expression constructs, expression vectors and recombinant  
 CC cells comprising pyranose oxidase nucleic acid sequences, and the  
 CC recombinant production of Trametes hirsuta pyranose oxidase. Pyranose  
 CC oxidase catalyses the oxidation of several monosaccharides in the  
 CC pyranose form at position C2 to produce 2-keto derivatives with the  
 CC release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta

CC pyranose oxidase may be used to produce the enzyme and to design  
 CC oligonucleotide probes to identify and clone genomic pyranose oxidase  
 CC cDNA or genomic DNA from different genera or species of microorganisms  
 CC (fungi or bacteria). The present sequence represents a cDNA encoding  
 CC pyranose oxidase from the fungus Trametes hirsuta.  
 CC Note: Both AAC87518 and AAC87519 are Trametes hirsuta pyranose oxidase  
 CC cDNA sequences which contain the entire open reading frame (ORF).  
 CC However, the two sequences have different stop codons - AAC87518 has an  
 CC opal stop codon while AAC87519 has an amber stop codon.  
 XX  
 SQ Sequence 1869 BP; 397 A; 622 C; 548 G; 302 T; 0 other;

Query Match 10.1%; Score 212.4; DB 22; Length 1869;  
 Best Local Similarity 51.5%; Pred. No. 2.2e-25;  
 Matches 841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;

QY 269 ggtacacacaagaagaatgaatcgaggttcacagaagaatattgacgcttcgcaatgta 328  
 Db 259 ggtcacacaagaagaacacccgtcgagtagcagaagaacatcgacaattcgtaattgt 318  
 QY 329 atcaaggagcccttacaacaagtctctgtcttcgtcagaaacacagacgtgcctacatt 388  
 Db 319 atacaagggaacattatgccgtctcgtgcccgtcaacacgatggtcgttgacagcta 378  
 QY 389 gatccggagcctggagcgcccccttgaaagttcagccatcgaacggtaaaatcct 448  
 Db 379 agccggcgctcatggcaagtttcgacg-----ttctgtccgcaacggggcgaaatcca 432  
 QY 449 caccagcgggaattcagaaactgtctggagggcgtgaacgctgagtcgcgccatg 508  
 Db 433 gagcaagaccgctgcgcacaccttagtgcagcggtcaccgcgtcgtcgcgccatg 492  
 QY 509 agtaccactggagctgctccacgcacaggtatcaccacccatggaagtctcccgggc 568  
 Db 493 tctacactggagctgcgacgcgcttcgagagcttcgagagctgcag-----537  
 QY 569 atcggccttcgagactcagtaacgacccggcgagagcgagacaaagtggaacgagctt 628  
 Db 538 ---cgcccgtcgtcgtgaaagaacgactccaaggcgacgacgagtgaggacaggtc 594  
 QY 629 tattcagagcggagcgtctcattcgggaattccacaaaggaattcagcagtgcaattcgg 688  
 Db 595 tacaagaagcggagctgctactcaagacggcgacacccagttccgagtgatccgc 654  
 QY 689 cacaccttcttcgctcttgcagacgctcaacaggtatcgtaacgtatcttcgc 748  
 Db 655 cacaactcgtgctcaagaagctgcaggaggagtagtaaaagg---gctgcgagcttcag 711  
 QY 749 cctctcccgctggcatgccaccggttgaaagaacgcgccgaatacgtcgaaatggcactca 808  
 Db 712 cagatcccgctcgcggcgacgcgcag-----gcccagcgttcgctgagtgagctcg 765  
 QY 809 gcagaaatctttccactctatctacaacgatgacaagcagagaagctcttaccctg 868  
 Db 766 gcgcacacgctgtt-----cgatctcgaagaccgcgcaacaaaggagcg 810  
 QY 869 ctgacgaacatcgtgcacacgactggcggttacggcggttatgagaagaagattggc 928  
 Db 811 ccgaagcagcgttcaacctcttcccgctcgtgcacgacgtgagggcgagataac 870  
 QY 929 gctgcgagtgaggaaatctactggccacacaggaatcctagtctcgagctggacagctat 988  
 Db 871 gcgaactcggagatcgtaggccttgatgtccgcacctccacgggggcaagacatcacc 930  
 QY 989 ataatggcaaggtatattgactgctggcggtgagcgatcggaacccacagattctctat 1048  
 Db 931 atcaagccaaggtgtacatcctcaccgcggcggttcacacacgcgcagctcctcgcg 990  
 QY 1049 aactcggggtt-----ctctgggtacaggtcagccacgcaatgac--tcgttgatc 1099  
 Db 991 gcctctggattcggcgagctgggtctcccgaccgcccaagccgctgcgtctctgctg 1050



[illegible][illegible]

RESULT 4  
AAT34420  
ID AAT34420 standard; cDNA; 1869 BP.

(KIKK ) KIKKOMAN CORP.

Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;  
Suzuki M;

WPI; 1996-278990/29.  
P-PSDB: AAR99628.

DNA encoding protein with pyranose oxidase activity at neutral pH - useful for the determination of glucose in body fluids or foods, or 1,5-anhydro-p-sorbitol used as marker for diabetes diagnosis

Claim 1: Page 10-13; 22pp; German.

The present sequence encodes a protein isolated from *Corioliolus versicolor*, which has the enzyme activity of pyruvate oxidase (PO). The PO oxidises glucose to gluconate and has an optimum pH of 7-7.5. It has a mol. wt. of 290000 (determined by gel filtration) and is stable at around 50deg.C. PO can be used for measurement of glucose in, e.g. foods or body fluids, or 1,5-anhydro-D-sorbitol which is an important marker used in the diagnosis of diabetes.

Sequence 1869 BP: 408 A: 601 C: 521 G: 339 T: 0 other: 0

Query Match 9.9%; Score 208.6; DB 17; Length 1869;  
Best Local Similarity 50.4%; Pred. No. 8.7e-25;  
Matches 856; Conservative 0; Mismatches 714; Indels 120; Gaps 0.

[illegible]



CC The invention relates to a novel antitumour protein extracted from  
CC fruiting bodies of the fungus pleurotus cornuopiae. The protein and  
CC nucleotides encoding it are useful for the treatment of cancer,  
CC including those caused by abnormal expression of tumour suppressor  
CC genes such as p53 and pBR. This sequence represents the coding  
CC sequence of the gene encoding the antitumour protein.  
CC  
CC  
XX  
SQ Sequence 1902 BP; 437 A; 556 C; 479 G; 430 T; 0 other;

Query Match	8.3%	Score 175.6;	DB 21;	Length 1902;
Best Local Similarity	57.1%	Pred. No. 1.5e-19;		
Matches 363;	Conservative	0;	Mismatches 264;	Indels 9; Gaps 2;

	Qy	1205	tgttgaaagaagcgttgtctcaacataattgcagaagaaccgacagatgycactgcccaatt	1264
	Db	1249	tgttgaacgaaggctccagaagcatagtatgcagcaccaggagatccccctcccgatc	1308
	Qy	1265	ccgttcgcgatccggaaccccgattacaaccccattitacagaagaacacccctggcac	1324
	Db	1309	ccgtcgatgacctgagccaagctcagcagcactttccaggacacacatccatggcac	1368
	Qy	1325	acgcagattcacccgcgatcttttctgtagtgcgtgcgtcctgagtggtgactctcgt	1384
	Db	1369	actcagattcatcgtagcctttagctacgcgcgtgtggcggagagtatcgacagccgt	1428
	Qy	1385	gtcatcgtcgacctgcgcgtgttttggcgcaacgcgacctgaagcaaacacacttttggtt	1444
	Db	1429	cttgtttgtactggcgtctcttcgtgcactgaacctgtggaggagaacaagctgtgg	1488
	Qy	1445	tctcagaacgatgttcaagacaggttacagtatgcgcgacgcagcgttcagatatcgg--a	1501
	Db	1489	tctcaaaacagatacctgatgcatacaacctgcccacagcccaactcagcttccggttc	1548
	Qy	1502	cccagcactgcgtccaaactgtagagcaaggaaaaatgatggccgatatatgtcgaaagtggcg	1561
	Db	1549	cccaggcgcacagcccaggaggcggagctcatgtatggctgacatgtgcacgatgtca	1608
	Qy	1562	agcaacttggaggttatitgccacgctccccccgcagtttatgtatccaggccttgca	1621
	Db	1609	accaagtcggtgtgctctgcggttcatatccgcagtttatgtgcccctggccttgta	1668
	Qy	1622	cttcacttgcggggactactgcgactggcttgcctgcagaaggcaactcacagtgc-----t	1675
	Db	1669	ctccacttggtgaaacccacgcataggcctttgatggaagctgataaaggctgtgtc	1728
	Qy	1676	gataacaactcgtgtgtctggacttggcatactttatgttgaggcaatggcaccaatc	1735
	Db	1729	gaccacaactcaaagttcttcgcatggagaacctgtccttggcgttgcggcgcaacatt	1788
	Qy	1736	aggacgggcttcggcgagAACCCGCACTTACGTTCGATGTGCCAGCTCATCAAGAGCGC	1795
	Db	1789	ggacgcgctatccttcgaacccgcgtcactgcggttgcgcttgcgaatcaggagctgt	1848
	Qy	1796	aggagcatcatcaatacactcaaggttgggactgac	1831
	Db	1849	aagtatatcaggaataatttcacgcgaactctctc	1884

## RESULT 6

AAV83626  
ID AAV83626 standard: cDNA to mRNA: 1701 BP.

AA AAV83626:

DT 26-FEB-1999 (first entry)

XX Nucleic acid encoding an antitumour protein.

Antitumour protein: Tricholoma matsutake: ss.

OS Tricholoma matsutake.

VV

Key	Location/Qualifiers
1..1701	
FT	/*tag= a
FT	/product= antitumour_protein
XX	
PN	JP10313876-A.
XX	
XX	
PD	02-DEC-1998.
XX	
PF	13-FEB-1998; 98JP-0031452.
XX	
PR	13-FEB-1997; 97JP-0029275.
XX	
PA	(MOMO-) MOMOYA KK.
XX	(NORQ ) NORINSUISANSHO SHOKUHN SOGO.
XX	
DR	WPI; 1999-074153/07.
DR	P-PSDE; AAW87531.
XX	
XX	
PT	An anti-tumour protein - prepared by culture of host cell
PT	transformed by vector containing base coding sequence
XX	
PS	Claim 3; Page 8-10; 15pp; Japanese.
XX	
CC	The present sequence encodes an antitumour protein, and i
CC	from Tricholoma matsutake.
XX	
SQ	Sequence 1701 BP; 411 A; 463 C; 432 G; 395 T; 0 other;

Query Match 7.2%; Score 152.6; DB 20; Length 1701;  
Best Local Similarity 53.5%; Pred. No. 7e-16;  
Matches 394; Conservative 0; Mismatches 304; Indels 39; Gaps 2;

QY 1106 ctgaggaggtacatcacggagcagccgatggcattttgccagatatagctcttgagggcagaaa 1165

Db 964 cttggtcgtacctaaccggaacagtcctgaactttttgtcagatcgttctcaagagggc 1023

Qv 1166 ttctcgacagcgtgcgcgacgattccttatggactgccattggtgaaagaagccattgct 1225

Db 1024 atagtcgatgccatcgctactgaccct-----cgcttcgctacgaaagttgaag 1071

Qv 1226 caacatatatcccaagaacccqacagatgcactgccattccgttccgcgattccggaaccc 1285

Db 1072 acgcacaaagaagacaccccgatgacgtactgccattccattccacgagcctgaacct 1131

Ov 1286 caggtaaacaaccccatTTacagaagaacaccccctggcacacgcagattcaccgcgatgct 1345

Db 1132 caagtgaattccgtacacgtcggacttcccttggcatgttcaggtccatccgcatcca 1191

Qy 1346 ttttcgtacggtaccggtcggtcctgaaggtgaactctcgtatcatcctcgaacctcgcctga 1405

[illegible]

0v	1406	tttgggcaaccggaccccttgaagcaaacacacctttagcttctc-----	1447
----	------	--	------

1252

1253

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

1212

1. *Chlorophyll a* (Chl *a*)  
 2. *Chlorophyll b* (Chl *b*)  
 3. *Chlorophyll c* (Chl *c*)  
 4. *Chlorophyll d* (Chl *d*)  
 5. *Chlorophyll e* (Chl *e*)  
 6. *Chlorophyll f* (Chl *f*)  
 7. *Chlorophyll g* (Chl *g*)  
 8. *Chlorophyll h* (Chl *h*)  
 9. *Chlorophyll i* (Chl *i*)  
 10. *Chlorophyll j* (Chl *j*)  
 11. *Chlorophyll k* (Chl *k*)  
 12. *Chlorophyll l* (Chl *l*)  
 13. *Chlorophyll m* (Chl *m*)  
 14. *Chlorophyll n* (Chl *n*)  
 15. *Chlorophyll o* (Chl *o*)  
 16. *Chlorophyll p* (Chl *p*)  
 17. *Chlorophyll q* (Chl *q*)  
 18. *Chlorophyll r* (Chl *r*)  
 19. *Chlorophyll s* (Chl *s*)  
 20. *Chlorophyll t* (Chl *t*)  
 21. *Chlorophyll u* (Chl *u*)  
 22. *Chlorophyll v* (Chl *v*)  
 23. *Chlorophyll w* (Chl *w*)  
 24. *Chlorophyll x* (Chl *x*)  
 25. *Chlorophyll y* (Chl *y*)  
 26. *Chlorophyll z* (Chl *z*)  
 27. *Chlorophyll aa* (Chl *aa*)  
 28. *Chlorophyll ab* (Chl *ab*)  
 29. *Chlorophyll ac* (Chl *ac*)  
 30. *Chlorophyll ad* (Chl *ad*)  
 31. *Chlorophyll ae* (Chl *ae*)  
 32. *Chlorophyll af* (Chl *af*)  
 33. *Chlorophyll ag* (Chl *ag*)  
 34. *Chlorophyll ah* (Chl *ah*)  
 35. *Chlorophyll ai* (Chl *ai*)  
 36. *Chlorophyll aj* (Chl *aj*)  
 37. *Chlorophyll ak* (Chl *ak*)  
 38. *Chlorophyll al* (Chl *al*)  
 39. *Chlorophyll am* (Chl *am*)  
 40. *Chlorophyll an* (Chl *an*)  
 41. *Chlorophyll ao* (Chl *ao*)  
 42. *Chlorophyll ap* (Chl *ap*)  
 43. *Chlorophyll aq* (Chl *aq*)  
 44. *Chlorophyll ar* (Chl *ar*)  
 45. *Chlorophyll as* (Chl *as*)  
 46. *Chlorophyll at* (Chl *at*)  
 47. *Chlorophyll au* (Chl *au*)  
 48. *Chlorophyll av* (Chl *av*)  
 49. *Chlorophyll aw* (Chl *aw*)  
 50. *Chlorophyll ax* (Chl *ax*)  
 51. *Chlorophyll ay* (Chl *ay*)  
 52. *Chlorophyll az* (Chl *az*)  
 53. *Chlorophyll aza* (Chl *aza*)  
 54. *Chlorophyll abz* (Chl *abz*)  
 55. *Chlorophyll acz* (Chl *acz*)  
 56. *Chlorophyll adz* (Chl *adz*)  
 57. *Chlorophyll aez* (Chl *aez*)  
 58. *Chlorophyll afz* (Chl *afz*)  
 59. *Chlorophyll agz* (Chl *agz*)  
 60. *Chlorophyll ahz* (Chl *ahz*)  
 61. *Chlorophyll aiz* (Chl *aiz*)  
 62. *Chlorophyll ajz* (Chl *ajz*)  
 63. *Chlorophyll akz* (Chl *akz*)  
 64. *Chlorophyll alz* (Chl *alz*)  
 65. *Chlorophyll amz* (Chl *amz*)  
 66. *Chlorophyll anz* (Chl *anz*)  
 67. *Chlorophyll aoz* (Chl *aoz*)  
 68. *Chlorophyll apz* (Chl *apz*)  
 69. *Chlorophyll aqz* (Chl *aqz*)  
 70. *Chlorophyll arz* (Chl *arz*)  
 71. *Chlorophyll asz* (Chl *asz*)  
 72. *Chlorophyll atz* (Chl *atz*)  
 73. *Chlorophyll auz* (Chl *auz*)  
 74. *Chlorophyll avz* (Chl *avz*)  
 75. *Chlorophyll awz* (Chl *awz*)  
 76. *Chlorophyll axz* (Chl *axz*)  
 77. *Chlorophyll ayz* (Chl *ayz*)  
 78. *Chlorophyll ayz* (Chl *ayz*)  
 79. *Chlorophyll azz* (Chl *azz*)  
 80. *Chlorophyll azaa* (Chl *aza*)  
 81. *Chlorophyll abz* (Chl *abz*)  
 82. *Chlorophyll acz* (Chl *acz*)  
 83. *Chlorophyll adz* (Chl *adz*)  
 84. *Chlorophyll aez* (Chl *aez*)  
 85. *Chlorophyll afz* (Chl *afz*)  
 86. *Chlorophyll agz* (Chl *agz*)  
 87. *Chlorophyll ahz* (Chl *ahz*)  
 88. *Chlorophyll aiz* (Chl *aiz*)  
 89. *Chlorophyll ajz* (Chl *ajz*)  
 90. *Chlorophyll akz* (Chl *akz*)  
 91. *Chlorophyll alz* (Chl *alz*)  
 92. *Chlorophyll amz* (Chl *amz*)  
 93. *Chlorophyll anz* (Chl *anz*)  
 94. *Chlorophyll aoz* (Chl *aoz*)  
 95. *Chlorophyll apz* (Chl *apz*)  
 96. *Chlorophyll aqz* (Chl *aqz*)  
 97. *Chlorophyll arz* (Chl *arz*)  
 98. *Chlorophyll asz* (Chl *asz*)  
 99. *Chlorophyll atz* (Chl *atz*)  
 100. *Chlorophyll auz* (Chl *auz*)  
 101. *Chlorophyll avz* (Chl *avz*)  
 102. *Chlorophyll awz* (Chl *awz*)  
 103. *Chlorophyll axz* (Chl *axz*)  
 104. *Chlorophyll ayz* (Chl *ayz*)  
 105. *Chlorophyll ayz* (Chl *ayz*)  
 106. *Chlorophyll ayz* (Chl *ayz*)  
 107. *Chlorophyll ayz* (Chl *ayz*)  
 108. *Chlorophyll ayz* (Chl *ayz*)  
 109. *Chlorophyll ayz* (Chl *ayz*)  
 110. *Chlorophyll ayz* (Chl *ayz*)  
 111. *Chlorophyll ayz* (Chl *ayz*)  
 112. *Chlorophyll ayz* (Chl *ayz*)  
 113. *Chlorophyll ayz* (Chl *ayz*)  
 114. *Chlorophyll ayz* (Chl *ayz*)  
 115. *Chlorophyll ayz* (Chl *ayz*)  
 116. *Chlorophyll ayz* (Chl *ayz*)  
 117. *Chlorophyll ayz* (Chl *ayz*)  
 118. *Chlorophyll ayz* (Chl *ayz*)  
 119. *Chlorophyll ayz* (Chl *ayz*)  
 120. *Chlorophyll ayz* (Chl *ayz*)  
 121. *Chlorophyll ayz* (Chl *ayz*)  
 122. *Chlorophyll ayz* (Chl *ayz*)  
 123. *Chlorophyll ayz* (Chl *ayz*)  
 124. *Chlorophyll ayz* (Chl *ayz*)  
 125. *Chlorophyll ayz* (Chl *ayz*)  
 126. *Chlorophyll ayz* (Chl *ayz*)  
 127. *Chlorophyll ayz* (Chl *ayz*)  
 128. *Chlorophyll ayz* (Chl *ayz*)  
 129. *Chlorophyll ayz* (Chl *ayz*)  
 130. *Chlorophyll ayz* (Chl *ayz*)  
 131. *Chlorophyll ayz* (Chl *ayz*)  
 132. *Chlorophyll ayz* (Chl *ayz*

.....

[illegible]

1471

[illegible]

**THE UNIVERSITY OF CHICAGO**





```
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 17-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH ) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2001-602746/68.
DR P-PSDB; AAU29061.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 2; Fig 75; 774pp; English.
XX
XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 4639 BP; 1426 A; 956 C; 1025 G; 1232 T; 0 other;
XX
XX
XX Query Match 4.5%; Score 94.4; DB 22; Length 4639;
XX Best Local Similarity 72.9%; Pred. No. 1.2e-06;
XX Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;
XX
XX Qy 1924 tctgtccctaccatgtgtgtagcagtagcggtgaaagatttgtgtattactga 1983
XX Db 4423 tatggctccattttattatagttaagattgtattccctaaagtgtgtttgtcga 4482
XX Qy 1984 acctgtactttgtctga----atagtttgccactatgattcatgtttataaaaaaaa 2038
XX Db 4483 cagtatcttttaaaagagctcttaaaaaataaggcatattgttcagttaaaaaaaa 4542
XX Qy 2039 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2098
XX Db 4543 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 4602
XX Qy 2099 aaaaaaa 2106
XX |||||||
```

```
Db 4603 aaaaaaa 4610
XX
XX RESULT 10
XX AAC78483
XX ID AAC78483 standard; cDNA; 4640 BP.
XX
XX AC AAC78483;
XX
XX DT 08-FEB-2001 (first entry)
XX
XX DE Human PRO708 (UNQ372) nucleotide sequence SEQ ID NO:113.
XX
XX KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
XX expressed sequence tag; detection; cancer; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200053756-A2.
XX
XX PD 14-SEP-2000.
XX
XX PF 18-FEB-2000; 2000WO-US04341.
XX
XX PR 08-MAR-1999; 99WO-US05028.
XX PR 12-MAR-1999; 99US-0123957.
XX PR 29-MAR-1999; 99US-0126773.
XX PR 21-APR-1999; 99US-0130232.
XX PR 28-APR-1999; 99US-0131445.
XX PR 14-MAY-1999; 99US-0134287.
XX PR 23-JUN-1999; 99US-0141037.
XX PR 26-JUL-1999; 99US-0145698.
XX PR 29-OCT-1999; 99US-0162506.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 02-DEC-1999; 99WO-US28551.
XX PR 02-DEC-1999; 99WO-US28565.
XX PR 16-DEC-1999; 99WO-US30095.
XX PR 30-DEC-1999; 99WO-US31243.
XX PR 30-DEC-1999; 99WO-US31274.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PR 06-JAN-2000; 2000WO-US00277.
XX PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
XX Goddard A, Godowski PJ, Grimaldi C, Gurney AL, Hillan KJ;
XX Kijavini IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
XX Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI; 2000-611443/58.
XX P-PSDB; AAB44257.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
XX to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX
XX Claim 2; Fig 42; 636pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX the presence of PRO polypeptides in samples, for linking bioactive
XX molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX can be used to kill the target cells, e.g. for the treatment of cancers.
XX The polypeptide pairs provide specific targeting of bioactive molecules
XX to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 4640 BP; 1427 A; 955 C; 1026 G; 1232 T; 0 other;
```











```
XX WPI: 2000-638216/61.
DR P-PSDB; AAB00196.
XX
XX Screening drug candidates for their ability to modulate breast cancer
PT by contacting the drug to a cell expressing an expression profile gene
PT and determining modulation of expression of the gene
XX
XX Disclosure; Fig 68; 258pp; English.
XX
XX New methods for screening drug candidates are described which
CC comprise adding a drug candidate to a cell that expresses a protein
CC selected from BCL1, BCL2, BCL7, BCL1, BCN5, BCQ2, BCQ5, BCR2, BCX2
CC and BCY3 or their fragments and determining the effect of the drug
CC on the expression of those proteins. Antibodies to breast cancer
CC genes (specifically BCL1 or its fragment (BCL1p1 or BCL1p2)) are
CC useful for inhibiting and treating breast cancer in individuals who
CC are non-responsive to anti-oestrogen and positive for oestrogen
CC receptor. Compositions comprising BCL1 or a nucleic acid encoding
CC BCL1 are useful for eliciting an immune response in an individual.
CC The antibodies are also useful for the diagnosis and prognosis of
CC breast cancer and for screening compositions which modulate the
CC breast cancer phenotype. The method allows rapid and simple
CC detection of lymph node metastases.
XX
XX Sequence 2710 BP; 851 A; 489 C; 525 G; 845 T; 0 other;
```

```
Query Match 4.3%; Score 91.2; DB 21; Length 2710;
Best Local Similarity 82.0%; Pred. No. 4 1e-06;
Matches 105; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1978 tactgaacctgtacttctgtctgaatgtatggtcactatgattcattgttaaaaaaaa 2037
   ||| | ||| ||||| || ||| || ||| ||||| |||||
Db 2583 tactcattgttaattgttaacacgtatcattataaacatttttgataaaaaaaa 2642

QY 2038 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2097
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2643 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2702

QY 2098 aaaaaaaa 2105
   ||||| |||
Db 2703 aaaaaaaa 2710
```

Search completed: September 3, 2002, 15:55:42  
Job time: 4746 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 13:18:27 ; Search time 2627.81 Seconds  
(without alignments)  
16771.108 Million cell updates/sec

Title: US-09-856-327-1  
Perfect score: 2106  
Sequence: 1 atcagcatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	2106	100.0	2106	6	BD006657	BD006657	A novel p	
2	212.4	10.1	1869	6	ARI41573	ARI41573	Sequence	
3	210.2	10.0	1869	6	I82410	Sequence 1		
4	210.2	10.0	1869	23	E11766	E11766	cDNA of pyr	
5	210.2	10.0	2046	8	D73369	D73369	Coriolus ve	
6	175.6	8.3	1902	6	E32737	E32737	Pennis cornu	
7	152.6	7.2	1701	6	ARI70485	ARI70485	Sequence	
8	129	6.1	1946	6	E33665	E33665	Novel prote	
9	129	6.1	1946	6	E33739	E33739	Reagent for	
10	94.6	4.5	2013	9	AB070201	AB070201	Macaca fa	
11	92	4.4	2360	10	BC008547	BC008547	Mus muscu	
12	92	4.4	2578	9	AB049758	AB049758	Homo sapi	
13	91.6	4.3	1618	3	AY069072	AY069072	Drosophil	
14	91	4.3	1282	9	BC017745	BC017745	Homo sapi	
15	90.8	4.3	1327	6	AX306545	AX306545	Sequence	
16	90.6	4.3	1567	9	AB047887	AB047887	Macaca fa	
17	90.6	4.3	2537	9	BC003684	BC003684	Homo sapi	
18	90.4	4.3	2335	9	AB055368	AB055368	Macaca fa	
19	90	4.3	2384	3	AY069709	AY069709	Drosophil	
20	89.8	4.3	470	3	AF177908	AF177908	Hydra vul	
21	89.8	4.3	1864	9	AB056421	AB056421	Macaca fa	
22	89.8	4.3	2180	8	AB028665	AB028665	Gentiana	
23	89.6	4.3	2032	9	AK027204	AK027204	Homo sapi	
24	89.4	4.2	2408	9	BC006525	BC006525	Homo sapi	
25	89.4	4.2	3295	10	BC005469	BC005469	Mus muscu	
26	89.2	4.2	1483	9	HSM800068	AL049283	Homo sapi	
27	89.2	4.2	3482	9	HSM800550	AL050393	Homo sapi	
28	88.8	4.2	2161	9	HSM801760	AL136792	Homo sapi	
29	88.6	4.2	1024	9	BC012602	BC012602	Homo sapi	
30	88.4	4.2	1091	6	AX053122	AX053122	Sequence	
31	88.4	4.2	5130	9	HSM801717	AL136749	Homo sapi	
32	88.2	4.2	1071	9	AF078844	AF078844	Homo sapi	
33	88.2	4.2	3364	8	ATH011628	AJ011628	Arabidops	
C	34	88	4.2	425	6	AX284646	AX284646	Sequence
35	88	4.2	1883	10	BC022165	BC022165	Mus muscu	
36	88	4.2	2003	9	AK000083	AK000083	Homo sapi	
37	88	4.2	2246	10	AF017152	AF017152	Mus muscu	
38	88	4.2	2390	9	AF090900	AF090900	Homo sapi	
C	39	87.8	4.2	310	6	AX187085	AX187085	Sequence
40	87.8	4.2	876	5	AF175983	AF175983	Rana sylv	
41	87.8	4.2	3353	6	AX086952	AX086952	Sequence	
42	87.8	4.2	3353	9	HSM801860	AL136892	Homo sapi	
43	87.6	4.2	2220	9	BC016137	BC016137	Homo sapi	
C	44	87.6	4.2	10039	6	AX346944	AX346944	Sequence
C	45	87.4	4.2	545	6	AX185705	AX185705	Sequence

ALIGNMENTS

RESULT 1

BD006657	BD006657	2106 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	A novel protein, a gene coding therefor and a method of using the same.				
DEFINITION	BD006657				
ACCESSION	BD006657.1 GI:18635028				
VERSION	JP 03075321-T/1.				
KEYWORDS	Lyophyllum shimeji.				
SOURCE	Lyophyllum shimeji.				
ORGANISM	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Lyophyllum.				
REFERENCE	1 (bases 1 to 2106)				
AUTHORS	Takakura, Y., Kuwata, S. and Inoue, Y.				
TITLE	A novel protein, a gene coding therefor and a method of using the same.				
JOURNAL	Patent: JP 03075321-T 1/16-FEB-2001;				
	JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHNO INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBUO CHO FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO INOUE				
COMMENT	OS Lyophyllum shimeji				
	PN JP 03075321-T/1				
	PD 16-FEB-2001				

PF 20-SEP-2000 JP 2000006404  
PR 21-SEP-1999 JP 99P 267238  
PI YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO INOUE PC  
C07K14/375, C12N15/31, C12N15/63, C12N1/21, C12Q1/68, C12P21/02, PC  
A01N65/00  
CC  
FH Key Location/Qualifiers  
FT CDS (8). (1861).  
1. 2106  
/organism="Lyophyllum shimeji"  
/db\_xref="taxon:47721"  
BASE COUNT 593 a 540 c 527 g 446 t  
ORIGIN

Query Match 100.0%; Score 2106; DB 6; Length 2106;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atcagccatgtctctcaacgagcagatgtacgagactatcacggtctatgcaaat 60  
|||||  
Db 1 ATCAGCCATGTCTCTCAACGAGCAGATGCTACGCGACTATCCACGGTCTATGCAAT 60  
|||||

Qy 61 caacggacagattcctcaagaacgcaattcacaagaacatacagaacgagagttgatgt 120  
|||||  
Db 61 CAACGGACAGATTCTCTAGAACGCAATTCACGAACATACGGAACACGCGAGTTGATGT 120  
|||||

Qy 121 attcattgagatcttgaccattggagcagctatgcaaatctctgtgtgaactgg 180  
|||||  
Db 121 ATTCAATGACGAGTCTGGACCCATTGGAGCGACGTATGCAAGCTCTGTGTGAAGCTGG 180  
|||||

Qy 181 tctacgtgtgtgagtcgagatcgagctgtgattcttctacgtgttaatgccga 240  
|||||  
Db 181 TCTACGTTGTGTGATGTCGAGATCGGAGCTGCTGTAGCTTCTACGCTGTATATGCCGA 240  
|||||

Qy 241 agaaggaaactcagttccctcattctctggtacacaagaagaatgaatcgagttcca 300  
|||||  
Db 241 AGAAGAACTCGATTCTCCCTACGTTCTCTGGGTACCAAGAAGAATGAATCGAGTTCCA 300  
|||||

Qy 301 gaaagattgacgccttcgaatgattcaaggagccttacaagaactctctgttcc 360  
|||||  
Db 301 GAAAGATTGACCGCTTCGTCGAATGAATCAAGGAGCCTTACAAAGTCTCTGTTC 360  
|||||

Qy 361 tgcagaacacagacgtgcctacattgatcccgagcctggagcgcgccccctggaag 420  
|||||  
Db 361 TGTCAAGAACAGAACGTCCTACATTGATCCCGAGCCTGGAGCGCCCCCTGGAG 420  
|||||

Qy 421 ttcagccatatacgaacggtgaaatcctcaccagcgggaaatcgagaaacttctgtgga 480  
|||||  
Db 421 TTCAGCCATATCGAACGTTAAAAATCTCAACAGCGGGAATTCGAGAACTTGTCTGCGGA 480  
|||||

Qy 481 ggcgtaacgctgagtcggcgcatagatcccacttgacgtgctccaagccacggaat 540  
|||||  
Db 481 GGCGTAAACGCTGGATGCGCGCATAGTACCCACTTGGACGTGCTCCAGCCACGGAT 540  
|||||

Qy 541 tcattccaccatgaaagctcccgccatcgccgctccgaagctcagtaacgaccggc 600  
|||||  
Db 541 TCATCCACCCATGGAAAGTCTCCCGGCATCGGCCGTCCGAGCTCAGTACAGCCCGC 600  
|||||

Qy 601 agagacgaaagagtggaacgagcttattccgagggcagagctcattcagggacttc 660  
|||||  
Db 601 AGAGACGAAAGAGTGGAAACGAGCTTATTTCGAGGCCGAGGCTCTCATCGGACTTC 660  
|||||

Qy 661 caccaggaattcgaagatcaattcggcacacctgttctgctcttttgaagacgc 720  
|||||  
Db 661 CACCAAGAAATTCACGAGTCAATTCGGCACACCCCTTGTTCGCGCTCTTTCGAAGACGC 720  
|||||

Qy 721 gtacaagatcgtaacgattcttctccctctccgttgccatggccaccgggttgaagaa 780  
|||||  
Db 721 GTACAGGATCGTCAACGATATCTTTCGCCCTCTCCCGTTCGATGCCACCGGTTGAAGAA 780  
|||||

Qy 781 cgcgcgggaaatcgtcgaaatggcactcagcagaataatcttttccactctatctacaaga 840  
|||||

Db 781 CGCGCGGAATACGTTCGAATGGCACTACGAGAAAAATCTTTTCCACTCTATCTACAAGGA 840  
|||||

Qy 841 tgaacagagaagaagctctttaccctgctacgaaccatcgctgacacacactggcct 900  
|||||

Db 841 TGACAAGCAGAAGAAGCTCTTTTACCTGCTGACGAACCATCGCTGCACACACTGGCGCT 900  
|||||

Qy 901 tacgggggggtatgagaagaagattggcgctgcgaggtcaggaatctactggccaccag 960  
|||||

Db 901 TACGGCGGGTATGAGAAGAGATTGGCGCTGCCGAGGTCAAGAACTACTTGGCCACCAG 960  
|||||

Qy 961 gaactactagttcgcagctggacagctatcatcatggcgaaggtatgtactggcgtggg 1020  
|||||

Db 961 GAATCCTAGTTCCGAGCTGGACAGCTATCATGCGGAAGGTATATGTACTTGGCGTCGGG 1020  
|||||

Qy 1021 agcatcggaacccacagattctctataactcgggtctctctctctggtctacaggtcacgc 1080  
|||||

Db 1021 AGCATCGGAACCCACAGATTCTCTATACTCGGGCTTCTCTGGGCTACAGGTACAGCC 1080  
|||||

Qy 1081 acgcaatgactcgttgatcccccaacctggggaggtacatcacggagcagcgatggcaat 1140  
|||||

Db 1081 ACGCAATGACTCGTTGATCCCCAACCTGGGGAGGTACATCACGGAGCAGCGATGGCAAT 1140  
|||||

Qy 1141 ttgcagatagttcgtgagcaggaattcgtcgcagcagcgtgcgcagcattccttggact 1200  
|||||

Db 1141 TTGCCAGATAGTCTTGAGGCAGGAATTCGTCGACAGCGTGGCGGACGATCCTTTATGGACT 1200  
|||||

Qy 1201 gccatgtggaagaacccgttctcaacatattgcgaagaacccagcagatgcactgcc 1260  
|||||

Db 1201 GCCATGTGGAAAGAACCCGTTGCTCAACATATTGCCAAGAACCCGACAGATGCACATGCC 1260  
|||||

Qy 1261 catccgtctccgcatcccgaaaccccggttaacaccccaattacagaagaacacccctg 1320  
|||||

Db 1261 CATTCGGTTCGCGCATCCGGAACCCAGGTAAACACCCCATTTACAGAAAGAACACCCCTG 1320  
|||||

Qy 1321 gcaacgcagattcacccgcgatgttttctcgtatggctgcgtcgtcgtcgtgaggtgactc 1380  
|||||

Db 1321 GCACACGAGATTCAACCGCATGCTTTTTCATACGGTCCGCTCGGCTCCTGAGGTGGACTC 1380  
|||||

Qy 1381 tctgtcatcgtcacctcgctggttctggcgcaacgacccctgaagcaaacacccctttt 1440  
|||||

Db 1381 TCGTGTATCATCTGACCTGCGCTGTTTGGCGCAACCGACCCCTGAAGCAACAAACCTTTT 1440  
|||||

Qy 1441 ggtttccagaaacgatgttcagacgggtacagtatgctgcgcagcgcagcttcagatatcg 1500  
|||||

Db 1441 GGTTTCCAGAACGATGTTCAGACGGGTACAGTATGCCGACGCGCGCTTCAGATATCG 1500  
|||||

Qy 1501 acccagcaactcgctcaaacacgtgagagcaagaaatgatggccgatatgtcgaaagtggc 1560  
|||||

Db 1501 ACCCAGCACTCGCTAAACGCTGAGAGCAAGGAAATGATGGCCGATATGTGCGAAGTGGC 1560  
|||||

Qy 1561 gacaaacttggaggttatttggccacgtcccccagtttctggtatccagggccttgc 1620  
|||||

Db 1561 GAGCAACTTGGAGGTATTATTTGCCACGTCCTCCCGCAGTTTATFGGATTCAGGCTTGC 1620  
|||||

Qy 1621 acttcaatttgcgggactactcgcattggcttcgacaaggaacactacagtggtcgataa 1680  
|||||

Db 1621 ACTTCACTTCGCGGACTACTCGCATTTGGCTTCGACAAGGCAACTACAGTGGCTGATAA 1680  
|||||

Qy 1681 caagtctgctggacttggcaatcttattgttgcagggcaatggcaccatacaggaac 1740  
|||||

Db 1681 CAAGTCGCTGCTGGACTTTTGGCAATCTTTATGTTTCAGGCAATGCACCATCAGGAC 1740  
|||||

Qy 1741 ggtcttcggcagaaacccgacattcagctcgtatgtccacgtctcaagagcagagag 1800  
|||||

Db 1741 GGGCTTCGGCAGAACCCGACACTTACGTCGATGTGCCACGCTATCAAGAGCGCGAGGAG 1800  
|||||

Qy 1801 catcataatacactcaagggtgggactgacggaaaaataacagcgagcgtcgcaacct 1860  
|||||

Db 1801 CATCATCAATACACTCAAGGTGGGACTGACGGAAAAAATAACAGCGGAGCATCGCAACT 1860  
|||||

Qy 1861 ttgagggaaggcacaacgagcagtgtaaaaaaacgcgtcaagtggcttacttcaagttgaatg 1920  
|||||





Db 1705 CCGCTCTTGGGCTTCAAGAACCTCTTCCTCGCGGCTCGGGAACATCCCAACCGCGTAC 1764  
Qy 1748 ggcgaagaccgacacttaccgtcgatgtgccacgctatcaagagcgcgagggagcatctc 1807  
Db 1765 GCGGGAACCGGACGCTCACCGCATGTCGCTTGGGATCAAGAGTGGGAGTACATCAAG 1824  
Qy 1808 aatacactcaag 1819  
Db 1825 AAGAACTTCGAG 1836

RESULT 3  
LOCUS I82410 1869 bp DNA linear PAT 10-JUN-1998  
DEFINITION Sequence 1 from patent US 5712139.  
ACCESSION I82410  
VERSION I82410.1 GI:3210707  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1869)  
AUTHORS Nishimura, I., Okada, K., Minamihara, T., Kawai, G., Koyama, Y. and Suzuki, M.  
TITLE Pyranose oxidase, pyranose oxidase gene, novel recombinant DNA and process for producing pyranose oxidase  
JOURNAL Patent: US 5712139-A 1 27-JAN-1998;  
FEATURES Location/Qualifiers  
source 1..1869  
BASE COUNT 408 a 593 c 529 g 339 t  
ORIGIN

Query Match 10.0%; Score 210.2; DB 6; Length 1869;  
Best Local Similarity 50.4%; Pred. No. 6.3e-33;  
Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;

Qy 200 gagatcgagctgtgtagcttctacgctgttaatgccgaaggaactgcagttccc 259  
Db 190 GAGCTCGTGAAGCGGGTTAAGGTGCCCATGTCGACATCGGGGAATTTGACTCTGGC 249  
Qy 260 tacgtctctgctaccacaagaagaatgaatcaggttccagaagaattgaccgcttc 319  
Db 250 CTGAAGATCGGTGCGCCACCAAGAAGAACACCGTCGAATACCAAGAACATTTGACAAGTTT 309  
Qy 320 gtcaatgtaatacgaaggagcgttacaacaagtctctgttctctgtcagaagaaccagaacgtg 379  
Db 310 GTGAACGTCTATCAGGGGCCAATTGATGTCTGTTTCCGTTCCCGTCAATACCCCTCGTGATC 369  
Qy 380 cctacactgtatcccgagcgctggagcgcgcgcctcctggaagttcagccatatacgaacggt 439  
Db 370 GACAGCTCAGCCCGACGCTCTGGCAAGCTTCATCG-----TTCCTCGTCGCAATGGC 423  
Qy 440 aaaaatcctaccagcggaattcagaaactgtctgtcgagggcgtaacgcgtggagtc 499  
Db 424 TCGAACCCAGAGCAGACCGCGCTTCGTAACCTCAGTGTGTCAGGGCGTCAACGCTGTCGTC 483  
Qy 500 ggcgcatgataccactgacgtgtctccagccacgaggtatcaccacctggaagagt 559  
Db 484 GGAGCATGTCCAGCATGTGACATGCGCGACACCGCGCTTTGA----- 527  
Qy 560 ctcccgagcatcgccgtccgaagctcagtaacgacccgcgagaggaacaaagagtgg 619  
Db 528 --CCGAGAGAGCGCCGCTTCTGCTGTAAGGACGACACGAGCGCTGACAGCGCGAGTGG 585  
Qy 620 aacgagctttattccgagggcgcgagcgtctcatcgcgggaacttcaccaagaattcagcag 679  
Db 586 CACCGGCTGTACACCAAGCGCGAGTCACTACTTCAAGACCGGACCGACCAAGTTCGAAGGAG 645  
Qy 680 tcaattggcaacaccttgctcgcgtctcttgcaagaacgcgtacaaaggaatcgtaacagt 739  
Db 1702 ACGGATCGCGCGGTGTTGGCTTCAAGAACCTGTTCCTCGGTGGTGGGAAACATTTCCC 1761

Db 646 TGGATCCGCCCAAAACCTCGTGCTCAACAAGCTCGCGAGGAATACAAAGGTGACGCGGAC 705  
Qy 740 atctttgccctctccgcttggaatgcccacgcggttgaagaacgcgcgcgaatacatcgtcaa 799  
Db 706 TTCCAGAGATCCCGCTCGCGCAACGCGTGC-----CAGTCCGACCTTCGTCGAG 756  
Qy 800 tggcactcagcagaaaaatcttttccactctatctacaacagatgacaagcagaagaagctc 859  
Db 757 TGGAGCTCGGCGAACAACCGTGT-----CGACCTCCAGAAGCGCGGAAC 801  
Qy 860 tttaacctgtgacgaacacatcgctgcacacgactggcgcttaacggcggtatgagaag 919  
Db 802 ACGGACGCGCGGAATGAGCGCTTCAACCTTCCCGCGGTTGCAATGTGAGCGCTCGTG 861  
Qy 920 aagattggcgctgcgaggtcaggaatctactggccaccaggaatccttagttcgcagctg 979  
Db 862 CGCAACACGTGAACTCCGAGATCGAGAGTCTGCAATCCAGGACCTCATCTCGGCGGAC 921  
Qy 980 gacagctatatacgtggcaaggtatgtactggcggtggagcgtacggaacccacag 1039  
Db 922 CGCTTCGAAATCAAAAGCAGACGTGTTCGTTCTTACAGCGCGGGCGGTCCACAACGCGCAG 981  
Qy 1040 attctctataactcgggctt-----ctctgggtacaggttcacgcccacgaatgactcg 1093  
Db 982 CTTCTCGTGAATCTGCGCTTTGGACAGCTGGCGCGCGGACCCCGGAACCCCGCGCAG 1041  
Qy 1094 ttgaccccaacctggggaggtacatcacgagcgcgagcgtatttggcagatagtc 1153  
Db 1042 TTGCTGCGCTCCCTGGGAAGCTACATCACGAGCAGTCCGCTGCTTCTGCGACAGCCGTG 1101  
Qy 1154 ttgagggcaggaattcgtcgacagcgtg-----cgcgacgactccttatgactgc----- 1203  
Db 1102 ATGAGCACCGAGCTCATCGACAGCGTCAAGTCCGACATGATCATCAGGGGCAACCTGGC 1161  
Qy 1204 -----atgg 1207  
Db 1162 GATCTGGGGTACAGCGTCACGTACACGCCCGCGCGGAGACCAACAAGACCCCGGACTGG 1221  
Qy 1208 tggaaagaagccgtgtctcaacatatcccaagaacccgcagatgcactgccattccg 1267  
Db 1222 TGGAAAGAAAGGTGAAGAACCACATGATGCAGACACGAGGAGGCCCTTCCAAATCCCG 1281  
Qy 1268 ttccgcgtacgcggaaccccgaggttaacaaccccatcttacagaagaacccccctggcacacg 1327  
Db 1282 TTCGAGGACCCCGAGCGCAGGTCAACCACTTGTTCAGGCCATCGCACCCCGTGGCACACT 1341  
Qy 1328 cagattcaaccgcgtgttttctgtaacggtgcgcgtcgtcctgaggttggaactcgtgctc 1387  
Db 1342 CAGATTACCCGCGATGCGTTTCAGTTTACGGCGCGGTGCAAGCAAGCATCGACTCAGGTCTC 1401  
Qy 1388 atcgtcgacctgcgctggtttggcgcaacccgacccctgaagcaaacacacttttggtttc 1447  
Db 1402 ATGCTGACTGGCGTTCCTCGCGCGAGCGAGCCAAAGGAGGAGAAACAAGCTCTGGTTC 1461  
Qy 1448 cagaacgagtgttaagacgggttacagttacgcgcagccagcttcagatatcgacccacg 1507  
Db 1462 TCGGACAAAATTACGGACACGTACAACATCGCGAGCGCGAGCTTCGACTTCGCGTTCGCG 1521  
Qy 1508 actg---cgtcaacggtgagagcaagaaaatgatggccgatattgctggaagtgagcagc 1564  
Db 1522 CGGGCCCGCACGAGCAAGGAGGCGGAGGACATGATGACCGATATGTGCGTATGTGTCGCG 1581  
Qy 1565 aacttgaggagtatttgccacgctcccccccgagtttatggatccagggccttgcaatt 1624  
Db 1582 AAGATTGGTGGCTTCCYCCCGCGGTCCCTCCCGCAATTCATGGAGCCCGGCTTGTTCCTT 1641  
Qy 1625 catctcgggggaactactcgcattggcttcgac-----aaggcaactacagtggtgcat 1678  
Db 1642 CACCTCGGTGTACGACCGCATGGGCTTCGACGAGCAGGAGGACAAAGTGTGCGTCAAC 1701  
Qy 1679 aacaactcgtcgtctgagggaactttgcaaatcttatgtgaggcaatggcaccateagg 1738  
Db 1702 ACGGATCGCGCGGTGTTGGCTTCAAGAACCTGTTCCTCGGTGGTGGGAAACATTTCCC 1761



Qy 1328 cagattaccgcgagctgttttctgtaoagtgccgtgctgctgagtgagtgactctgctc 1387  
|||||  
Db 1342 CAGATTACCGCGATCGCTTTCAGTTCAGCGCGGTGCAGCAAGCATCGACTCAGCTCTC 1401  
Qy 1388 atcgtcagctgcgctggtttgggcaaccagcactgaagcaacaacacttttggtttc 1447  
|||||  
Db 1402 ATCGTCGACTGCGCTTCTTCGCGCGGACGAGCAAGAGGAGCAACAAAGCTCTGGTTC 1461  
Qy 1448 cagaacgagtggtcaagcgggtacagtgatccgcagccgacgttcagatcgcaccagc 1507  
|||||  
Db 1462 TCGGCAAAATACGACACGTACACATGCGCGAGCGGACGCTTCGACTTCGCTCCG 1521  
Qy 1508 actg---cgtcaacgtgagagcaaggaatgatggcggatgatgtgcgaagtggcgagc 1564  
|||||  
Db 1522 CGGGCCGACGAGCAAGAGCGGAGGACATGATGACCGATATGTGGCTTATGTGCGCG 1581  
Qy 1565 aacttggaggttatttgcacagctcccccgcagtttattgattcagcagccttgcaatt 1624  
|||||  
Db 1582 AAGATTGTGTGCTTCTGCGCGGCTCCCTCCGCAATTTCATGGAGCCGCTTGTGCTT 1641  
Qy 1625 catcttgcgggaactactcgcattggttcgac-----aagcgaactacagtggtgat 1678  
|||||  
Db 1642 CACTCGTGTGACGACCGATGGCTTCAGGACGAGGAGCAAGTGTGCTGCTCAAC 1701  
Qy 1679 aacaactcgtgttgggaatttgcgaatttattgttcgaggaatggcaccatcagg 1738  
|||||  
Db 1702 ACGGATCGCGGTGTTGGCTTCAAGAACCTGTCCTCGGTGGTGGCGGAACATTCCTC 1761  
Qy 1739 acgggcttcgagagaaacacacttacgtcgtatgtgcacgctatcaagagcgcagg 1798  
|||||  
Db 1762 ACCGGTACGGCGGAACCGACGCTACCGCAATGTGCTGCGGATCAAGAGTTGCGAG 1821  
Qy 1799 agcatcatcaactca 1817  
|||||  
Db 1822 TACATCAGACAACTTCA 1840

RESULT 5  
D73369  
LOCUS D73369 2046 bp mRNA linear PLN 04-FEB-1999  
DEFINITION Coriolus versicolor mRNA for pyranose oxidase, complete cds.  
ACCESSION D73369  
VERSION D73369.1 GI:1845548  
KEYWORDS pyranose oxidase; PROD.  
SOURCE Coriolus versicolor (strain:ps4a) cDNA to mRNA.  
ORGANISM Coriolus versicolor  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
Aphyllophorales; Coriolaceae; Coriolus.  
REFERENCE  
AUTHORS Nishimura,I.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-1995) Ikuko Nishimura, Kikkoman Corporation,  
Research and Development Division; Noda 399, Noda, Chiba 278, Japan  
(Tel:0471-23-5571, Fax:0471-23-5550)  
REFERENCE  
AUTHORS Nishimura,I., Okada,K. and Koyama,Y.  
TITLE Cloning, sequencing analysis, and expression in Escherichia coli of  
pyranose oxidase cDNA of Coriolus versicolor  
JOURNAL Unpublished (1995)  
REFERENCE  
AUTHORS Nishimura,I., Okada,K. and Koyama,Y.  
TITLE Cloning and expression of pyranose oxidase cDNA from Coriolus  
versicolor in Escherichia coli  
JOURNAL J. Biotechnol. 52 (1), 11-20 (1996)  
MEDLINE 97177816  
COMMENT Submitted (5-Oct-1995) to DDBJ by:  
Ikuko Nishimura  
Dept. of Research and Development Division  
Kikkoman Corporation  
Noda, Chiba 278  
Japan  
Phone: 0471-23-5571

FEATURES  
source 0471-23-5550.  
Location/Qualifiers  
1..2046  
/organism="Coriolus versicolor"  
/strain="ps4a"  
/db\_xref="taxon:57466"  
26..1897  
/note="PROD"  
/codon\_start=1  
/product="pyranose oxidase"  
/protein\_id="BAAL119.1"  
/db\_xref="GI:1845549"  
/translation="MSTSSDPFFNFNTKSSFRSAAQKASATSLPLPLGPKKVPKGM  
IKYDVIVGSGPIGCTYARELVKAMFDIGEIDSLKIGAHKNTVEYOKNIDK  
FNVIOQLMSVSPVNTLVIDTSLPTSWAOSFFVFRNGSNPDPLRNLSGOAVTRV  
VGNSTHTWCATPRFDEORPLLVKDDODADDAEDWRLYTKASYFKTCTDQKESIR  
HNLVLKLAEEYKQDRFQIPLAATRRSTFVEMSSANTVFLQNRPTDAPNEREN  
LFAVACERVVNTNSIEISLHDLISGRFEIKADVFVLTAGVHNAQLLVNSGF  
GOLGRPDNPOLLPSLSYITEQSLVFCQTMSTELIDSVKSDMIRNGPGLDYS  
VTYTPGAETNKHNDWNEKVNMMQHOEDPLPIPFEDPEQVTTLPQSPHPTQTH  
RDAPSYGAVOQSIDSLRIVDRFRFTEPEKEENKLPFSDKITDTYNNPQTFDFRPA  
GRPSKEAEDMNTDMCVMSAKIGGFLPGSLPQFMPEGLVLHGGTHRMPDEQDKCV  
NTDSRVFGFKNLFLGGCGNPTATGANTLTATSLAKSCEYIKNNFTSPFPFTDQAE"  
2046  
polya\_site 450 a 634 c 575 g 387 t  
BASE COUNT  
ORIGIN  
Query Match 10.08; Score 210.2; DB 8; Length 2046;  
Best Local Similarity 50.4%; Pred. No. 6.4e-33;  
Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;  
Qy 200 gagatcgagctgctgatagcttctacgctgttaatgcggaagaagcagcttc 259  
|||||  
Db 215 GAGTCGTCGAGCGGTTACAGTCCCAATGTCGACATCGGGAAATGACTCTGGC 274  
Qy 260 tacgttcctggctaccacaagaagaatgaaatcagcttccagaagaatgacgcgttc 319  
|||||  
Db 275 CTGAAGATCGGTGCCACAGAAGAACACCGTCGAATACCAAGAACATTTGACAAGTTT 334  
Qy 320 gtcagttaatacaggagccttacacaagtctctctgttcgtcagaacacagcgtg 379  
|||||  
Db 335 GTGAACGTCATTACAGGGCCAAATGATGCTCTTCCGTTCCCGTCAATACCTCGTGATC 394  
Qy 380 cctacacttgatccggagcgtgagcgcccccctcgaagtccagccatcgaacgtt 439  
|||||  
Db 395 GACAGCTACCCGCGAGCTCTTGGCAAGCTTCATCG-----TTCCTCGTCCGCAATGGC 448  
Qy 440 aaaaatcctcaccagcggaattcagaaacttgctgcggaggccgtaacgcgtggagtc 499  
|||||  
Db 449 TCGAACCAGAGCAGGACCCGCTTCGTAACCTCAGTGTGTCAGGCGGTACCGGTGTCGTC 508  
Qy 500 ggcggcatgataccactgagctgctccacgcagcaggtattcaccaccatggaagt 559  
|||||  
Db 509 GGAGGCATGTCCACGCACTGGACATGCGGACACCGCGCTTTGA----- 552  
Qy 560 ctccgggcatcgccctgcgaagctcagtaacacccgcgagagcagacaaagatgg 619  
|||||  
Db 553 --CCGCGAGCAGCCCGCTTCTGCTGAGGACACACAGGAGCTGACCGCGGATGGC 610  
Qy 620 aacgagctttattccgagcgagcgtctcattcagggactccaccagaagattcgaagc 679  
|||||  
Db 611 GACCGGTGTACACCAAGCGCGAGTCACTTCAAGACCGGACCGGACCGACTTCAAGGAG 670  
Qy 680 tcaattcggaacacccctgttctgcgtctcttgcgaagcgcgctacagagatcgtaacgt 739  
|||||  
Db 671 TCGATCCGCCACAACTCTGCTCAACAAGCTCGCGGAGGAATACAAAGGTACAGCGGAC 730  
Qy 740 atcttgcctctctccgttggcattgcccacggttggaagaacgcgcggaatacgtcaa 799  
|||||  
Db 731 TTCCAGCAGATCCCGCTCGCGGCAACCGGTCTG-----CAGTCCGACCTTCGTCGAG 781  
Qy 800 tggcactcagcagaaatcttttccactctatataacagatgacaagcagaagaagctc 859

```
Db 782 TGGAGCTCGGGGAACACCGGTGTT-----CGACCTCCAGAACAGGCCCAAC 826
QY 860 ttaccctgtgacgaaccatcgtgcacacgactggtcttacggtcggcgggtatgagaag 919
Db 827 ACGGACGCGCGGAATGATGACGCTTCAACCTCTTCCCGCGGTGCATGTGAGCGGTGCTG 886
QY 920 aagattggcgtccgaggtcgaagaatctactggtccacccaggaatctagtgtgcagctg 979
Db 887 CGCAACACGTCGAATCCGAGATCGAGAGTCTGCACATCCACGACCTCATCTCGGGCGAC 946
QY 980 gacgctatatcgtgcgaaggtatatgtactggtcgtcggagcgtatcgcaaacacag 1039
Db 947 CGCTTCAATCAAGCAGAGTGTGCTTACAGCCGGCGGTGCCAACAGCGCGAC 1006
QY 1040 attctctataactcgggctt-----ctctgggtctacaggtcagccacgcaatgactcg 1093
Db 1007 CTCTCTGTGAATCTGTGGCTTTGGACAGCTGGCGCGCGGACCCCGGGAACCCGCGCGAC 1066
-QY 1094 ttgattcccaactcgggaggtacatcacaggagcagcgtatggtcatttttccagatagtc 1153
Db 1067 TTGCTGCGCTCCCTGGGAAGCTATATCACCAGCAGTCTGCTGCTTCTTCCAGACCGGTG 1126
QY 1154 ttgaggcaggaattcgtcgacagctg-----cgcgacgactccttatgtgactgcc----- 1203
Db 1127 ATGAGCACCAGCTCATCGACGCTCAAGTCCGACATGATCATFCAGGGGCAACCCCTGGC 1186
QY 1204 -----atgg 1207
Db 1187 GATCTGGGTACAGCGTCACGTACACGCCGCGCGGAGACCAACAGCACC CGACTGG 1246
QY 1208 tgaagaagaagcgttgtctcaacatatgtccaagaaccgcagatgcactgcccattccg 1267
Db 1247 TGGAAACGAAAGGTGAAGAACCCATCATGATGACGACCAGCAGGAGACCCGCTTCCAATCCG 1306
QY 1268 ttccgcatcgaaacccaggttaaacacccatttacagaagaacacccctggcacag 1327
Db 1307 TTCGAGACCCCGAGCGCGAGGTACACCTTGTTCAGCCATCGCACACC CGTGGCACACT 1366
QY 1328 cagattcacgcgtatgttttttgcaggtgcgctgctggtcgtgaggtggactctcgtg 1387
Db 1367 CAGATTACCCGCGATGGTTTCAGTTACGGCGCGGTGCAGCAAAAGCATCGACTCAGCTCTC 1426
QY 1388 atcgtgcacctgcgtggttttgcgaacacgacccctgaagaacaaacaccttttggtttc 1447
Db 1427 ATCGTGCACCTGGCGCTTCTTCGCGCGAGCGAGGCCAAAGGAGGAGAAAGCTCTGGTTC 1486
QY 1448 cagaacgatgttcaaacggttacagtatccgacgcgacgttcagatcgaaccagc 1507
Db 1487 TCGGACAAAATTACGGACACGTACAACATCCCGACCGCGACGTTTCGACTTCCCGTCCCG 1546
QY 1508 actg---cgtcaaacgtgagagcaaggaataatgatggccgatatgtgcgaagtggcgagc 1564
Db 1547 GCGGGCGCAGCAGCAAGAGGCGGAGGACATGATGACCCGATATGTGCGTTATGTGCGCG 1606
QY 1565 aacttggaggttatttggccacgctcccccgcagtttatgtatgcagcccttgcaatt 1624
Db 1607 AAGATTGGTGGCTTCCTGCCCGCTCCCTCCGCAATTATTCGAGGCCGCTCTTGCTCTT 1666
QY 1625 catcttgcgggactactcgcattgtgcttcgac-----aaggcaactacagtggctgat 1678
Db 1667 CACTTCGGTGTACGACCCGATGGGCTTCGACGAGCAGGAGGACAAAGTCTCGCTCAAC 1726
QY 1679 aacaactcgtgtctgggaggttggccaatctttatgttgcaggcaatggcaccatcagg 1738
Db 1727 ACGGACTCGGCGCTTTGGCTTCAAGAACCTGTCTCTCGGTGGCTCGCGAAACATTCCT 1786
QY 1739 acggttctggcgagaaacccgacacttaoctgcgtatgtgcacgctatcaagagcgaggy 1798
Db 1787 ACCGCGTACGCGGCAACCCGAGCTCACGGCAATGTGCGTCTCGGATCAAGAGTTGGGAG 1846
QY 1799 agcatcatcaataactca 1817
||||| ||| |||
```

```
Db 1847 TACATCAAGAACAACACTTCA 1865
RESULT 6
LOCUS E32737 1902 bp DNA linear PAT 07-FEB-2001
DEFINITION Panus cornucopeiae-derived antitumor protein and gene thereof.
ACCESSION E32737
VERSION E32737.1 GI:13017467
KEYWORDS JP 1999315096-A/1.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1902)
AUTHORS Akihiro,M.K.I.I. and Kido,T.S.R.R.
TITILE Panus cornucopeiae-derived antitumor protein and gene thereof
JOURNAL Patent: JP 1999315096-A 1 16-NOV-1999;
NYUFUDO KURIESHON GIJUTSUKENKYUKUMIAI
COMMENT OS Pleurotus cornucopeiae
PN JP 1999315096-A/1
PD 16-NOV-1999
PF 07-AUG-1998 JP 1998236349
PR
PI AKIHIRO MORITA, KOJI IZUMO, KAZUAKI KIDO, TOMOHIDE SAKA, PI
RYOKO AOKI
PC C07K14/375, A61K35/84, A61K38/00, C07K16/14
CC
FH
FT
FEATURES
source Location/Qualifiers
BASE COUNT 437 a 556 c 479 g 430 t
ORIGIN
Query Match 8.3%; Score 175.6; DB 6; Length 1902;
Best Local Similarity 57.1%; Pred. No. 7.5e-26;
Matches 363; Conservative 0; Mismatches 264; Indels 9; Gaps 2;
QY 1205 tgggtgaagaagccgttgctcaacatatgtccaaagaaccgcagatgcactgcccatt 1264
Db 1249 TGGTGAACAGAGAAGGTCCAGAAAGCATGATGCAGCAGCAGGAGGATCCCTCCCGATC 1308
QY 1265 ccgttcgcgcatcggaaacccaggttaaacacccatttacagaagaacacccctggcac 1324
Db 1309 CCCTCGATCCCTTGAGCCACAGGTCCACGACTCTTCCAGGACACACATCCATGGCAC 1368
QY 1325 acgagattcacgcgcatgcttttgcaggtgcgctgcgtgcgtgaggtggactctcgt 1384
Db 1369 ACTCAGATTTCATGTCGTATGCTTTAGCTACGGCGCTGTGGCGGAGAGTATCGACAGCCGT 1428
QY 1385 gtcatcgtgcacctgcgtggttttgcgaacacgacccctgaagaacaaacaccttttggtt 1444
Db 1429 CTTGTTGTTGACTGGCGCTTCTTCGGTTCGCACTGAACCTGTGGAGGAGAAACAAGCTGG 1488
QY 1445 ttccagaacgatgttcaagcgggtacagatgccagcgcgacgttcagatatcg---a 1501
Db 1489 TTCTCAAAAACAGATCACTGATGATCAACTGCCACAGCCCACTTCAGCTTCCCGTTC 1548
QY 1502 cccagcactcgttcaaacgtgagagcaaggaataatgatggccgatatgtgcgaagtggcg 1561
Db 1549 CCCAGGGCGGCACAGCCAGGAGCGGAGCTCATGATGCTGCATGTGTCACAGATGCA 1608
QY 1562 agcaattgggaggttatttggccacgctcccccgcagtttatgtatgcacggccttgca 1621
Db 1609 ACCAAAGTCGGTGGCTTCCTGCCCGTTCATATCCGCGTTCATCCGAGTTTATGGCTCCTGCCCTGTA 1668
QY 1622 cticattctgcgggactactcgcattgtgcttcgacaaggaactacagtggc-----t 1675
Db 1669 CTCACCTTGGTGGAAACCCCAACCGCATGGGCTTTGATGAGGAAGCTGATATAAGCGGTGTGTC 1728
```



[illegible]

FEATURES  
source Location/Qualifiers

1..1946  
/organism="Tricholoma matsutake"  
/db\_xref="taxon:40145" 471 t

BASE COUNT 484 a 516 c 475 g

ORIGIN

Query Match 6.1%; Score 129; DB 6; Length 1946;  
Best Local Similarity 52.5%; Pred. No. 2.5e+16;  
Matches 387; Conservative 0; Mismatches 305; Indels 45; Gaps 3;

Qy 1106 ctgggaggtacatcacgcagcagccgatggcatctttccagatagtcttgaggcgaggaa 1165  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1064 CTGGTCGCTACCTCAGCGAACAGTCCATGACTTTTGTCAGATGTGTTCTCAAGAGGAGC 1123  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1156 ttgttcgacagcgtgctgcgcagcagctacttattggaactgcattggtgaaaagaagccgttgtct 1225  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1124 ATAGTCGAATTCCATCGCTTACTGACCT-----CGCTTGCCTGCCAAGGTTTGAG 1171  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1236 caaatattgccagaaccgcagacagatgcactgccattcccgttccgcgtaccgaagcccc 1285  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1172 CGGCACAGAAGACAGCCCCGATGACGTGTCGGATTCATTCCACGAGCGTGAACCT 1231  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1286 caggttaaacaccccatattacagaagaacacccctggcacacagcagattcacccgatgct 1345  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1232 CAAGTGATGATTCCGTACACGTGCGACTTCCTTTGGCATGTTTCAGGTCATCGCTATGCA 1291  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1346 ttttctgtagtgtccgtcgtgctctgaggtgacctctdgtcaatctgtcgacctgcgtcg 1405  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1292 TTT-----GGTGATGTTGGACCCAAAGCCGACCOCGTGTGTGTCGTGATCTGAGGTTT 1345  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1406 ttggcgcaaccgcctggaagcaacaacctttggttttcc----- 1448  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1346 TTCGCCAATCAGATATTGTGCGAAGAAATCGAGTACTTTCGGTCCCAACCTTAAGCTA 1405  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1449 -----agaacgatgtttaagacgggttacagtatccgcgcagccgacgtttccagatat 1498  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1406 CGCGACTGGGAAGCGGGTGTTCAGACACACTTATGGAATGCCACGCCACATTCATGTC 1465  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1499 cgaccagcactcgtcaaacgtgagagcaagaaaaatgatggccgatgtgcgaatg 1558  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1466 AAGCGGACCAACGCCGATGGAGACCGTGACCAGAGGATGATAATGATGACCAACGTC 1525  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1559 gcgagcaacttggagggttatttggccacgtcccccccgagtttatgatgccaggcctt 1618  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1526 CGGAACATACTGGCGGGTACTTCTTGGCTCTCACCTCAATTTATGGCACCTGGTCTC 1585  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1619 gaacttcaattcgtgggagactactgcattggtctgcagaaggaaccaatacagtgagctgat 1678  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1586 GCACAGCACATCAGCGGAACTACTCGGATCGGACAGATGATCAAACTTCTGTGCTGAT 1645  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1679 acaactcgtcgttcgggactttgccaattttatgttcaggagcaatggcaccatcagg 1738  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1646 CCGACATCAAAGTTCATAACTTCGACAATCTGTGGGTCTGGCGGGAATGGGTGCATTCOA 1705  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1739 acgggcttcggcgagaacccacacttactcgtatgtgcccgctatcaagagcgcgagg 1798  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1706 GATCGACTGCTGCAACCCGACTCGTACGAGCGTGGCGTATGCGCTTAAGGGTGTCTGAG 1765  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1799 agcatcatcaatacaact 1815  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1766 GCTGTAGTCAAGTTACCT 1782  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10  
AB070201  
LOCUS AB070201 2013 bp mRNA linear PRI 16-AUG-2001  
DEFINITION Macaca fascicularis testis cDNA clone:QtsA-17226, full insert  
sequence.  
ACCESSION AB070201  
VERSION AB070201.1 GI:15208242





QY 2079 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106  
|||||  
Db 2330 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2357  
|||||

RESULT 12  
AB049758  
LOCUS  
DEFINITION Homo sapiens mawbp mRNA for MAWD binding protein, complete cds. PRI 23-AUG-2001  
ACCESSION AB049758  
VERSION AB049758.1 GI:10800085  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens liver cDNA to mRNA.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Iriyama,C., Matsuda,S., Katsumata,R. and Hamaguchi,M.  
TITLE Cloning and sequencing of a novel human gene which encodes a putative hydroxylase  
JOURNAL Journal of human genetics. 46 (5), 289-292 (2001)  
MEDLINE 21253940  
REFERENCE  
AUTHORS Matsuda,S. and Iriyama,C.  
TITLE Direct Submission  
JOURNAL Submitted (10-OCT-2000) Satoru Matsuda, Nagoya University School of Medicine, Molecular Pathogenesis; Tsurumai 65, Showa-ku, Nagoya 466-8550, Japan (E-mail:smatsuda@med.nagoya-u.ac.jp, Tel:81-52-744-2463, Fax:81-52-744-2464)

FEATURES  
source  
1. .2578  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="liver"  
165..932  
/gene="mawbp"  
165..932  
/gene="mawbp"  
/codon\_start=1  
/product="MAWD binding protein"  
/protein\_id="BAB16606.1"  
/db\_xref="GI:10800086"  
/translation="MHQIARENLSETAFIRKLHPTDNPFASSCGFLRWFTPASEVP  
LCGHATLASAAVLHFKIKNNSTLTFTLSELARRAEDGIVLDPLYPAPQDFHE  
VEDLITAGNTLVQDICYSPDTKLLVRLSDYNRSLFNKLVNTENLQVNTGKV  
KGLILTKGPGGQTOAFDYGRYPAPWGVADPVTGSAHAVLSSYWSQHLGKKEMH  
AFQCSHRGGELGILSRPDGRVDIRGGAUVLGTLT"  
2453  
polyA\_site  
/note="120 a nucleotides"  
BASE COUNT 875 a 477 c 544 g 682 t  
ORIGIN

Query Match 4.4%; Score 92; DB 9; Length 2578;  
Best Local Similarity 78.8%; Pred. No. 9.5e-09;  
Matches 110; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1967 attttgtattactgaacctgtactttgtctgaatagttatgacctatgattcatgtt 2026  
|||||  
Db 2393 ATTATTATAGCAGTATCTGTGTTCTCTGATGTGCAATAATCAATAATTCACGAG 2452  
|||||

QY 2027 taaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2086  
|||||  
Db 2453 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2512  
|||||

QY 2087 aaaaaaaaaaaaaaaaaaaaaa 2106  
|||||  
Db 2513 AAAAAAAAAAAAAAAAAAAAAA 2532  
|||||

RESULT 13  
AY069072

LOCUS  
DEFINITION Drosophila melanogaster GH09066 full length cDNA. INV 17-DEC-2001  
ACCESSION AY069072  
VERSION AY069072.1 GI:17861479  
KEYWORDS  
SOURCE  
ORGANISM Drosophila melanogaster  
fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS 1 (bases 1 to 1618)  
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,  
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,  
Nunco,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,  
Yu,C., Lewis,S.E., Rubin,G.M. and Celnikier,S.  
Direct Submission  
Submitted (10-DEC-2001) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our Web site  
(http://fruitfly.berkeley.edu) or send email to  
cdna@fruitfly.berkeley.edu.

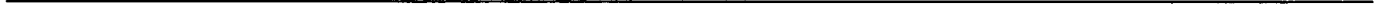
FEATURES  
Location/Qualifiers  
1. .1618  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="49E1-49E1"  
1. .1618  
/gene="CG17019"  
/note="alignment with genomic scaffold AE003820"  
/db\_xref="FLYBASE:FBgn0033783"  
291..1250  
/gene="CG17019"  
/note="Longest ORF"  
/codon\_start=1  
/db\_xref="FLYBASE:FBgn0033783"  
/product="GH09066p"  
/protein\_id="AAL39217.1"  
/db\_xref="GI:17861480"  
/translation="MGESTQALNVNORPANPSAFYPAVPORPKFKYTRRSQGLNRR  
YHSSDDESPVAPGCGCLSLGLSTLSEHPHSHRSCQRCQCGKNKINIRHVEKM  
RHLENSQSEEDIKRELQEFLLYLEQTKSVSDADSHAVSPLSDAISVAAGRS  
PMPITPTTFEFTSQDEDIRWDDDEGIHYAAPDFEPTAGIDSRVNLDFDLKDL  
EGILTKQLKELMLHRVDYKGCCKEQLDRLSRKMTKRECPAVEKPLATDELKICM  
DAPIECVFLECGHMACTSCGKVLNCPICRCQIVRVVRFRA"  
BASE COUNT 490 a 375 c 423 g 330 t  
ORIGIN

Query Match 4.3%; Score 91.6; DB 3; Length 1618;  
Best Local Similarity 63.8%; Pred. No. 1.1e-08;  
Matches 139; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1889 aaacgctcaagtgcctactcaagttgaatgcattctgtccctaccatgttgatgtg 1948  
|||||

[illegible]

Search completed: September 3, 2002, 15:49:51  
Job time: 9084 sec



**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 12:40:02 ; Search time 1745.12 Seconds  
(without alignments)  
16288.042 Million cell updates/sec

Title: US-09-856-327-1  
Perfect score: 2106  
Sequence: 1 atcagccatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	95.4	4.5	480	9 AI392214	AI392214 NCSP1F9T7
C 2	95.2	4.5	451	9 AL514867	AL514867 NCSP1F9T7
C 3	94.2	4.5	165	10 BI746603	BI746603 km29d01.Y
C 4	94.2	4.5	286	10 BG736597	BG736597 rk59f08.Y
C 5	93.8	4.5	318	10 BI941823	BI941823 sc80a12.Y
C 6	93.6	4.4	503	9 AL513809	AL513809 sh45g11.Y
C 7	93.4	4.4	162	9 AW395255	AW395255 sh45g11.Y
C 8	93.4	4.4	638	9 AL513901	AL513901 AL513901
C 9	93.2	4.4	484	10 BI680498	BI680498 458776 MA
C 10	93	4.4	600	10 BG926622	BG926622 HNC56-1-H
C 11	93	4.4	685	9 AV714975	AV714975 AV714975
C 12	92.6	4.4	262	10 BE722980	BE722980 192270 MA
C 13	92.6	4.4	337	10 BI142697	BI142697 kt44f06.Y
C 14	92.6	4.4	421	10 BG362036	BG362036 gb50f09.Y
C 15	92.2	4.4	308	10 BE633260	BE633260 uv74g09.Y
C 16	92.2	4.4	565	9 AL514205	AL514205 AL514205
C 17	92	4.4	175	10 BI748585	BI748585 ro71g08.Y

18	92	4.4	321	9 BE059229	BE059229 sn27h09.Y
19	92	4.4	363	10 BF612286	BF612286 df06h03.Y
20	92	4.4	367	10 BI715621	BI715621 ic35b05.Y
21	91.8	4.4	213	9 AI496256	AI496256 sb01e02.Y
22	91.6	4.3	266	10 BI782468	BI782468 kh26b06.Y
23	91.6	4.3	464	10 BI313535	BI313535 dad77d01.Y
24	91.6	4.3	510	10 BE632703	BE632703 uv64c06.Y
C 25	91.6	4.3	926	10 BM394635	BM394635 50072-2-5
26	91.4	4.3	200	10 C93770	C93770 C93770 D1ct
27	91.4	4.3	303	10 BG652798	BG652798 sad84c02.Y
C 28	91.4	4.3	368	9 AU176554	AU176554 AU176554
29	91.2	4.3	190	10 BM025324	BM025324 fs77a04.Y
30	91.2	4.3	383	10 BI784726	BI784726 saf9ze03.Y
31	91.2	4.3	474	10 BI840173	BI840173 fs70d08.Y
32	91.2	4.3	2091	11 AF111849	AF111849 Homo sapi
33	91	4.3	286	10 BG737121	BG737121 rk67a06.Y
34	91	4.3	327	9 BE023867	BE023867 sm93d12.Y
35	91	4.3	342	9 AW156527	AW156527 sc28c06.Y
36	91	4.3	344	10 BI946237	BI946237 sc71b05.Y
37	91	4.3	356	10 BM367068	BM367068 NXLV_044
38	91	4.3	519	10 BM283661	BM283661 K150e03.Y
39	90.8	4.3	312	9 AW317448	AW317448 sg49h10.Y
40	90.8	4.3	363	10 BM130000	BM130000 Pb24h09.Y
41	90.8	4.3	603	10 BG925918	BG925918 HNC21-1-F
42	90.6	4.3	233	10 BI941734	BI941734 sc88e06.Y
43	90.6	4.3	296	9 BE057871	BE057871 sn08c11.Y
44	90.6	4.3	361	9 AU029579	AU029579 AU029579
C 45	90.6	4.3	456	9 AL513913	AL513913 AL513913

#### ALIGNMENTS

RESULT 1

AI392214/C

LOCUS

DEFINITION

3' similar to nmt1 protein homolog - thiamine biosynthesis enzyme,

AI392214

NCSP1F9T7 Subtracted Perithecial Neurospora crassa cDNA clone SP1F9

480 bp mRNA linear EST 03-FEB-1999

AI392214

EST

KEYWORDS

ORGANISM

Neurospora crassa

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariales; Sordariaceae; Neurospora

1 (bases 1 to 480)

Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,

Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E.,

Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R.,

Ortega, J., Pavlova, I., Perez, J., Todisco, S., Trujillo, R.,

Valentine, J., Wells, A., Werker-Washburne, M., Yazzie, S. and Natvig

D.O.

Expressed sequences from conidial, mycelial, and sexual stages of

Neurospora crassa

Fungal Genet. Biol. 24, 348-363 (1997)

97435549

Contact: Natvig, D.O./Nelson, M.A.

Department of Biology

University of New Mexico

Castetter Hall, Albuquerque, NM 87131, USA

Tel: 505 277 3411

Fax: 505 277 0304

Email: natvig@biology.unm.edu.

Location/Qualifiers

1. .480

/organism="Neurospora crassa"

/strain="fl a (FGSC 4347)"

/db\_xref="taxon:5141"

/clone="SP1F9"

/clone\_lib="Subtracted Perithecial"

/sex="Mating type a (fluffy), fertilized"

/tissue\_type="Perithecia"









(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 101 a 14 c 12 g 35 t  
ORIGIN

Query Match 4.4%; Score 93.4; DB 9; Length 162;  
Best Local Similarity 78.3%; Pred. No. 0.19; Indels 0; Gaps 0;  
Matches 112; Conservative 0; Mismatches 31;

QY 1964 aagatttctgtattactgaacctgtacttctgtctgaatagttatgacacatgattcat 2023  
DB 20 AATAGCTTGTTGTTTACTTCTATCCCATTTTACATGTTTGGATGGAGTGGTTATTCAT 79  
QY 2024 gtttaaaaaa 484 bp mRNA linear EST 13-FEB-2001  
LOCUS AL513901 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CL0BA006ZG01 3  
DEFINITION prime, mRNA sequence.  
ACCESSION AL513901  
VERSION AL513901.1 GI:12777395  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 638)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
Bp 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
source  
1..638  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CL0BA006ZG01"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/notes="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 347 a 58 c 15 g 141 t 77 others  
ORIGIN

Query Match 4.4%; Score 93.4; DB 9; Length 638;  
Best Local Similarity 60.8%; Pred. No. 0.099; Indels 0; Gaps 0;  
Matches 121; Conservative 16; Mismatches 62;

QY 1908 ttcaagtgtgaatgcattctgtgtccctaccattgtgtgtacgataggcgttgaaaga 1967  
DB 217 TTTTWTGTAATWATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 158  
QY 1968 ttitgtgtattactgaacctgtacttctgtctgaatagttatgacacatgattcatgttt 2027  
DB 157 TTTTWTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 98  
QY 2028 aaaaaa 484 bp mRNA linear EST 17-SEP-2001  
LOCUS BI680498 458776 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
DEFINITION BI680498  
ACCESSION BI680498  
VERSION BI680498.1 GI:15633409  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 484)  
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACACGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 127 row: L column: 3  
Seq primer: ATTAGTGACACTATAG.

FEATURES  
source  
1..484  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/notes="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 175 a 103 c 83 g 123 t  
ORIGIN

Query Match 4.4%; Score 93.2; DB 10; Length 484;  
Best Local Similarity 88.6%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 101; Conservative 0; Mismatches 13;

QY 1993 ttgtctgaatagttatggcactgattcatgttttaaaaaaaaaaaaaaaaaaaaaa 2052

```

Db 369 TTGCTTAACAAATTATGGAATGATACCTATTTTAAATAAAAAAAAAAAAAAAAAAAAA 428
Qy 2053 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
Db 429 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 482

RESULT 10
LOCUS BG926622
DEFINITION HNC56-1-17.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
sequence.
ACCESSION BG926622
VERSION BG926622.1 GI:14321145
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 600)
AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
TITLE Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthritic cartilage cDNA libraries
JOURNAL Osteoarthritis Cartilage 9 (7), 641-653 (2001)
MEDLINE 21482651
COMMENT Contact: Sanjay Kumar
UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-legsk.com
Seq primer: T7.
FEATURES
source
Location/Qualifiers
1..600
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HNC (Human Normal Cartilage)"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 230 a 101 c 112 g 157 t
ORIGIN

Query Match 4.4%; Score 93; DB 10; Length 600;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 120; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1942 tgatgtgacataggcgctgaagattttgtattactgaacctgtacttctgtgaa 2001
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 377 TGATTTATTTGGAGTGGAGCATGTTTTTAAAGAAAACATGTCATGTAGGTGTCTAAA 436

Qy 2002 tagttatggcactatgattcattgtttaaaaaaaaaaaaaaaaaaaaaaaaaa 2061
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 437 AATAAATGCTATTAACCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 496

Qy 2062 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 497 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 541

RESULT 11
LOCUS AV714975
DEFINITION AV714975 DCB Homo sapiens cDNA clone DCBAYA01 5', mRNA sequence.
ACCESSION AV714975
VERSION AV714975.1 GI:10796492

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 685)
AUTHORS Xu,X., Gu,J., Liu,F., Qu,J., Zhao,M., Li,Y., Huang,Q., Zhou,J.,
Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X.,
Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Cheng,Z. and Han,Z.
TITLE Homo sapiens cDNA DCB clones
JOURNAL unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
Location/Qualifiers
1..685
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DCBAYA01"
/clone_lib="DCB"
/cell_type="dendritic cells"
/dev_stage="mature"
/lab_host="BM25.8"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB"
BASE COUNT 271 a 118 c 123 g 173 t
ORIGIN

Query Match 4.4%; Score 93; DB 9; Length 685;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 120; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1942 tgatgtgacataggcgctgaagattttgtattactgaacctgtacttctgtgaa 2001
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 TGATTTATTTGGAGTGGAGCATGTTTTTAAAGAAAACATGTCATGTAGGTGTCTAAA 363

Qy 2002 tagttatggcactatgattcattgtttaaaaaaaaaaaaaaaaaaaaaaaaaa 2061
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 364 AATAAATGCTATTAACCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 423

Qy 2062 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 468

RESULT 12
LOCUS BE722980
DEFINITION 192270 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE722980
VERSION BE722980.1 GI:10124276
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 262)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnkruug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pertes,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 16:01:24 ; Search time 28.03 Seconds  
(without alignments)  
750.079 Million cell updates/sec

Title: US-09-856-327-2\_COPY\_76\_618

Perfect score: 2901

Sequence: 1 NAEGETAVPYVPGYHKKEI.....IINTLKGTDGKNTGEHRL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	116	4.0	514	1	G6PD_MYCTU
2	101	3.5	1010	1	SCAM_RICPA
3	100	3.4	1142	1	ENAM_PIG
4	98	3.4	925	1	W70T_HUMAN
5	97.5	3.4	639	1	AMYG_ASPAK
6	97.5	3.4	1536	1	SIN3_YEAST
7	97	3.3	612	1	AMYG_ASPOR
8	96	3.3	291	1	US02_HSV11
9	96	3.3	964	1	YIN0_YEAST
10	95.5	3.3	519	1	AMYL_SACFI
11	95.5	3.3	708	1	HELS_SULSO
12	95.5	3.3	1257	1	CCAA_BACTU
13	95	3.3	500	1	YDAK_YEAST
14	94.5	3.3	773	1	CDH_PHACH
15	94.5	3.3	855	1	GAFL_SCHPO
16	93.5	3.2	406	1	YNQ5_YEAST
17	93	3.2	470	1	LEU2_AZOVI
18	93	3.2	737	1	AMYL_AEDAE
19	93	3.2	890	1	GLND_ECOLI
20	93	3.2	890	1	GLND_SALTY
21	93	3.2	955	1	VP2_BTIV17
22	93	3.2	1012	1	UBAL_SCHPO
23	92	3.2	376	1	JBUC_MOUSE
24	92	3.2	704	1	RP3A_BOVIN
25	92	3.2	813	1	KRAF_CAEEL
26	92	3.2	1584	1	U104_CAEEL
27	92	3.2	1638	1	DPO3_LACLA
28	91.5	3.2	425	1	IUCD_ECOLI
29	91	3.1	993	1	YAV1_SCHPO
30	90.5	3.1	847	1	FAV3_HUMAN
31	90	3.1	463	1	GUN_BACSP
32	90	3.1	487	1	ENGA_CHLPN
33	90	3.1	492	1	FTSA_HELPY

#### ALIGNMENTS

RESULT 1

ID	G6PD_MYCTU	STANDARD;	PRT;	514 AA.
AC	O08407;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).			
GN	ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCY493.07.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RP	[.]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala E.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,			
RA	Hornaby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RN	Nature 393:537-544(1998).			
RP	[2]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-			
CC	1,5-lactone 6-phosphate + NADPH.			
CC	-1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.			
CC	-1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE			
CC	FAMILY.			
CC	-1- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE			
CC	A CLASSICAL ZWF.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; Z95844; CAB09259.1; -.			



OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Enamel epithelium;  
 RX MEDLINE=98040070; PubMed=9372788;  
 RA Hu C.-C., Fukae M., Uchida T., Oian Q., Zhang C.H., Ryu O.H.,  
 RA Tanabe T., Yamakoshi Y., Murakami C., Dohi N., Shimizu M.,  
 RA Shimer J.P.;  
 RT \*Cloning and characterization of porcine enamel mRNA.  
 RL J. Dent. Res. 76:1720-1729(1997).  
 RN [2]  
 RP SEQUENCE OF 39-773 FROM N.A., SEQUENCE OF 39-49; 174-276;  
 RP 515-574; 535-578; 641-646; 663-665; 670-686; 740-750; 765-773 AND  
 RP 833-848, AND CHARACTERIZATION.  
 RX MEDLINE=97350624; PubMed=9206327;  
 RA Fukae M., Tanabe T., Murakami D., Dohi N., Uchida T., Shimizu M.;  
 RT \*Primary structure of porcine 89 kDa enamel.  
 RL Adv. Dent. Res. 10:111-118(1996).  
 CC -1- FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS  
 CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE  
 CC COVERING VERTEBRATE TEETH.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.  
 CC INTACT ENAMELIN AND LARGE-MOLECULAR-WEIGHT ENAMELINS ARE LIMITED  
 CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,  
 CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER  
 CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CRYSTALLITES IN ROD  
 CC AND INTERROD ENAMEL.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE  
 CC TRANSITION STAGE.  
 CC -1- PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL  
 CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN  
 CC OCCURS SOON AFTER SECRETION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U52196; AAD10837.1; -  
 DR GlycoSuiteDB: O97939; -  
 KW Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.  
 FT SIGNAL 1 38  
 FT CHAIN 39 1142 ENAMELIN.  
 FT CHAIN 39 ? 56 KDA ENAMELIN.  
 FT CHAIN 39 665 89 KDA ENAMELIN.  
 FT CHAIN 39 ? 142 KDA ENAMELIN.  
 FT CHAIN 39 ? 155 KDA ENAMELIN.  
 FT CHAIN 174 276 32 KDA ENAMELIN.  
 FT CHAIN 515 665 25 KDA ENAMELIN.  
 FT CHAIN 670 ? 34 KDA ENAMELIN.  
 FT CHAIN ? ? 45 KDA ENAMELIN.  
 FT MOD\_RES 53 53 PHOSPHORYLATION (PROBABLE).  
 FT MOD\_RES 191 191 PHOSPHORYLATION.  
 FT MOD\_RES 216 216 PHOSPHORYLATION.  
 FT MOD\_RES 547 547 PHOSPHORYLATION.  
 FT CARBOHYD 245 245 HYDROXYLATION.  
 FT CARBOHYD 252 245 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 680 680 H -> D (IN REF. 2).  
 FT CONFLICT 838 840 RDH -> TTI (IN REF. 2).  
 SQ SEQUENCE 1142 AA; 128352 MW; 9383068C87CC5FC6 CRC64;

Query Match 3.4%; Score 100; DB 1; Length 1142;  
 Best Local Similarity 19.2%; Pred. No. 11;  
 Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;

QY 1 NAEETAVPYVP--GYH-----KKNEIEFKDIDR-----FVNVIK 35  
 DB 192 NEEGNN--PYGFGYHGFGGRPPYSEEMFEQDFEKPKDPKTKTETPATESVN--- 245  
 QY 36 ALQOVSVVRNQNVPYL-DP-GAUSAPPSSAISNGKNPHOR-----EFNLSEAVTR 87  
 DB 246 ----TTPETNSTQPNAPNPRGNDTSPGTG--GQGPNSRNPSTQNGPAPNVSGQGVPR 299  
 QY 88 ----GVGGMSTHWTCTPPI-----HPPMESLPGIGR---PKLSNDPAEDDKEW 129  
 DB 300 SQSPWGPQPIIHENYENPNIRGFPARRQWRPPG---PAMGHRNGPFYRMOQLQGRPW 356  
 QY 130 NELYSEARLIGTSTKEFDESIRHTLVLSLDQDAYKQRIFRPLPLACHRLKNAPEYVE 189  
 DB 357 NSFTLEGG-----QAVRPGYPTVRYVY-----GSTARSNPPNTA- 390  
 QY 190 WHSAENLFHSIYNDOKKOKLFTLLTNHRCPLALTG-----GYEKKIGNAE 235  
 DB 391 -GNSANLRRKPEGPNK-----NPMVTNVAAPPKPGKGTVDQDNENIQNPREKVSQKE 440  
 QY 236 VRNLLATRNPS-----SOLDSYIMAKVYVVLASGAIGNPOLLYNSGFGLOVTPRNDSLPN 291  
 DB 441 -RTVVPTRDPSGWRNSQDYGINKSNYKL-----PQ-----PENMLVFN 479  
 QY 292 LG-----RYITEQPMACQIVLRQEFVDSVDRDDPYGLPMWKEA---VAQ 332  
 DB 480 FNSIDQRENSYPRGESKRAPNSDGTQTQIIPK-----GIVLEPRRIPYSETNQPELK 534  
 QY 333 HIAKNP--TDALPTPRDPEP-----QVTPPTFEEHWHQIHRDAFSYGAVGPEV 381  
 DB 535 HSAYPQVYTGIPSPAKEHPFAGRNNTWNOQEIISPFKED-PGRQEHPLHLSHG----- 587  
 QY 382 DSRVIVDLRWFEGATDPEANLLVFQNDVQDYSMP-----QPTFRYRPSASTNVARKMA 437  
 DB 588 -SRVHYVPDYNDPNSYLSRNTWYERDSDPNTMGQPNENPHYMNTPPDET----- 642  
 QY 438 DMCEVASNLGGYLTSPQPMDFGLALHLAGTTTIGFDKATTVADNNSLVWDFANLYVAG 497  
 DB 643 -----IPYNEEDPIDPTGDEHFPGQSR-----WDMELSFKPE 674  
 QY 498 NGTIRTCFGENPT 510  
 DB 675 DPTVRHYEGEQYT 687  
 RESULT 4  
 W70T\_HUMAN STANDARD; PRT; 925 AA.  
 AC P57737;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE 70 kDa WD-repeat tumor rejection antigen homolog.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT \*NEDO human cDNA sequencing project.  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).  
 CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; AK025674; BAB15211.1; -;	
DR	InterPro: IPR001680; WD40.	
DR	Pfam: PF00400; WD40; 6.	
DR	PRINTS: PR00320; GPROTEINRPT.	
DR	SMART: SM00320; WD40; 6.	
DR	PROSITE; PS00678; WD_REPEATS_1; 1.	
DR	PROSITE; PS50082; WD_REPEATS_2; 4.	
DR	PROSITE; PS50294; WD_REPEATS_REGION; 2.	
KW	Repeat; WD repeat.	
FT	75	WD 1.
FT	REPEAT	WD 2.
FT	124	WD 3.
FT	REPEAT	WD 4.
FT	166	WD 5.
FT	REPEAT	WD 6.
FT	209	WD 7.
FT	REPEAT	WD 8.
FT	542	
FT	REPEAT	
FT	592	
FT	REPEAT	
FT	635	
FT	REPEAT	
FT	728	
FT	REPEAT	
FT	925 AA; 100574 MW; A393CE973C94F4A9	CRC64;
SQ	SEQUENCE	

Query Match 3.4%; Score 98; DB 1; Length 925;  
Best Local Similarity 20.2%; Pred. No. 12;  
Matches 88; Conservative 52; Mismatches 139; Indels 156; Gaps 21;

[illegible]

RESULT	5	
AMYG_ASPAK		
ID	AMYG_ASPAK	STANDARD; PRT; 639 AA.
AC	P231176;	
DT	01-NOV-1991	(Rel. 20, Created)
DT	01-NOV-1991	(Rel. 20, Last sequence update)
DT	16-OCT-2001	(Rel. 40, Last annotation update)
DE	Glucosylase I precursor (EC 3.2.1.3)	(Glucan 1,4-alpha-glucosidase)
DE	(1,4-alpha-D-glucan glucohydrolase).	
GN	GAI.	
OS	Aspergillus awamori (var. kawachi).	
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
OC	Eurotiates; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.	



```
FT CARBOHYD 514 514 O-LINKED (BY SIMILARITY).
FT CARBOHYD 516 516 O-LINKED (BY SIMILARITY).
FT CARBOHYD 517 517 O-LINKED (BY SIMILARITY).
FT CARBOHYD 519 519 O-LINKED (BY SIMILARITY).
FT CARBOHYD 521 521 O-LINKED (BY SIMILARITY).
FT CARBOHYD 523 523 O-LINKED (BY SIMILARITY).
FT CARBOHYD 524 524 O-LINKED (BY SIMILARITY).
FT CARBOHYD 525 525 O-LINKED (BY SIMILARITY).
FT CARBOHYD 526 526 O-LINKED (BY SIMILARITY).
FT CARBOHYD 527 527 O-LINKED (BY SIMILARITY).
FT CARBOHYD 528 528 O-LINKED (BY SIMILARITY).
FT CARBOHYD 529 529 O-LINKED (BY SIMILARITY).
FT CARBOHYD 530 530 O-LINKED (BY SIMILARITY).
FT CARBOHYD 531 531 O-LINKED (BY SIMILARITY).
FT CARBOHYD 533 533 O-LINKED (BY SIMILARITY).
FT CARBOHYD 534 534 O-LINKED (BY SIMILARITY).
SQ SEQUENCE 639 AA; 68271 MW; E112B31A4DD8D6B CRC64;

Query Match 3.4%; Score 97.5; DB 1; Length 639;
Best Local Similarity 20.2%; Pred. No. 7.4;
Matches 106; Conservative 51; Mismatches 194; Indels 175; Gaps 25;

QY 80 LSAEAVTRGVCGSTHTWCSTPRHHPMESLPGIGRPKLSNDPAEDDKENNELYSEAERL 139
DB 106 ISSQAIIGGVNSPGDL-----SSGGIGRKPKNVDETAYTGSGRPORDGPAL 153
QY 140 IGTSKTEFDESIRHTLVRLSLQADYKD-RQRIFRPLPLACHRLKNAPEYVE--WHSANL 196
DB 154 RAYAMIGFGQVL-----LDNGYTSAAEYIHWPL-----VRNDLSYVAQYWNQVG-- 197
QY 197 FHSYNDKOKKFLTLNHRCTRLAUTGG--YEKKTGAAFEVRLNLAIRNPSSQLDSYIM 254
DB 198 -YDLWEEVNGSSFTTIAVQHR---ALVEGSAFATAVGSS-----CSWCD50A- 240
QY 255 AKYVVLASGAIGNPOIL--YNSGFGLOVTPRNDLSLPN-----LGRYTEQPMARCO- 305
DB 241 -----POILCYLOSFMTGYILANFDSRRSGKDTNLLGSIHTFDPEAGCDD 287
QY 306 -----IVLROEVDVSR-----DDPY--GLPMWKEVAQ 332
DB 288 STQPQSPRALANHKVEVDSFRSYITLNDGLSDSEAVAGVGYEDSYNGNPFQSTLA- 346
QY 333 HIAKNPTDALPIPRDPPEQVTTFTTEHPWHTQ-----IHRDAFSYGAVG--- 378
DB 347 -AAEQLYDAL-----YQWDKQGSLEITDVSLEDFKALYSGAATGYS 387
QY 379 --PEVDSRVIVDLRWFG-----ATDPEANLLVFQNDVQDG--YSMPQPTFRYRSTA 427
DB 388 SSSSTYSIVSAVKTFADGFGVSVIVETHAASNGSLSEQFDKSDGDELSARDLTWSYAALLT 447
QY 428 SNVRARKM-----ADMCEVASNLGGY---LPTSPQFMDPGALHLAGTTRI 472
DB 448 ANNRNSVVPWSGCTSSASHPGTCATASGTVSSYVTSWPSIVATGCTTTTATTGGS 507
QY 473 G-----FDKATTVADNNSLVWDFAN-----LYVAGNCTIIRTFGCEN 508
DB 508 GGVTSKTTTTSKTTSTSTTSCTTPTAVAVTFDLTATTITYGEN 553

RESULT 6
ID SIN3_YEAST STANDARD; PRT; 1536 AA.
AC P22579; Q08049;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Paired amphipathic helix protein.
GN SIN3 OR SDI1 OR UME4 OR RPD1 OR GAM2 OR SDS16 OR YOL004W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
```

```

RN SEQUENCE FROM N.A.
RP STRAIN=GRF88;
RX MEDLINE=91042523; PubMed=2233725;
RA Wang H., Clark I., Nicholson P.R., Herskowitz I., Stillman D.J.;
RT "The Saccharomyces cerevisiae SIN3 gene, a negative regulator of HO,
RL contains four paired amphipathic helix motifs.";
RN Mol. Cell. Biol. 10:5927-5936(1990).
RN SEQUENCE FROM N.A.
RP Hughes B., Pohl T.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND
CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STE6,
CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SPO13), GENES
CC THAT RESPOND TO EXTERNAL SIGNALS (PHO5) AND TRK2. IT IS PROBABLY
CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR
CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3
CC -1- SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL
CC REGULATORY PROTEIN RPD3.
CC -1- SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.
CC -1- DOMAIN: CONTAINS 4 PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A
CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A
CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.
CC -1- SIMILARITY: TO S.POMBE SPAC12C2.10C.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M36822; AAA34839.1; -;
DR EMBL; Z74746; CAA99003.1; -;
DR PIR; S12068; RGBY53.
DR SGD; S0005364; SIN3.
DR InterPro; IPR003822; PAH.
DR Pfam; PF02671; PAH; 3.
KW Repeat; Transcription regulation; Repressor; Cell division;
KW Activator; Nuclear protein... 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.
FT DOMAIN 239 1200
FT REPEAT 239 252 HELIX A.
FT REPEAT 274 285 HELIX B.
FT REPEAT 426 439 HELIX A.
FT REPEAT 461 472 HELIX B.
FT REPEAT 679 692 HELIX A.
FT REPEAT 714 725 HELIX B.
FT REPEAT 1152 1165 HELIX A.
FT REPEAT 1189 1200 HELIX B.
FT DOMAIN 480 519 GLN-RICH.
FT CONFLICT 510 510 Q -> QAO (IN REF. 2).
SQ SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;

Query Match 3.4%; Score 97.5; DB 1; Length 1536;
Best Local Similarity 19.4%; Pred. No. 27;
Matches 103; Conservative 71; Mismatches 163; Indels 193; Gaps 28;

QY 58 SAPPGSAISGNKPNHOREFNLSA-----EAVTRGVCGMSTHTWCSTPRHHPMESLP 111
DB 1044 SGSDGSSIASRKRYPQOQESLLDILHRSRYQKLKR-----SNDEGKVPQSEPPPEEP 1098
QY 112 G-IGRPKLSNDPAEDD-----KENNELYSEA 136
DB 1099 NTIEEELIDEAEAKNPWLGNLVEANSQGIQNRISFNLFANTNIYIFFRHTTIV--- 1155
QY 137 ERLGTGTSTKFEDESIRHTLVRLSL-----QDAYKDRQIRFRP 173
DB 1156 ERLI--EIKQNNERTVTEINTRSVTFKADLLDLSQSEMGDFVGEDAYKQVLRSLR- 1212
QY 174 LPLACHRLKNAPEYVEVWHSANLPHSYNDKOKKFLTL-----LTNHRCTRLAALGGY 227
```

```

Db 1213 -----RLNGDLERHWF--EBSLRQAYN-NKAFKLYTIDKVTQSLVKHAHT--LMTDAK 1261
QY 228 EKKIGAAEVLNLIATNPSSQLDSYIMAKVYVVLASGAIGNPQILYNSGFGSLQVTPRND 287
Db 1262 TAEIMALFVKD---RN-----ASITSKADQIYR-----LQVR-SHMS 1295
QY 288 LIPNLGR-----YITQPMAFQOIVLRQEFVDSVRDDPYGLPMWKEAVAQHTAKNPT 339
Db 1296 NTENMFRIEDKRTLTHSSQYIALDDLTLEKPADEK-----WKYYVTSYALPHPT 1347
QY 340 DALPIFRPEQVTPPTTEHPWHQIHRDAFSYGAVGPEVDSRVIVDLRFEGATDPEA 399
Db 1348 EGI-----PHEKLIKPFLE-----RLIEFGQDIDGTEVDEEF-----SPEG 1383
QY 400 NNL-----LVFONDQDG-YSM--POPTFRYRPSTASNVNRKRMADMCEVASNLG 447
Db 1384 ISVSTLIKIKIQTIVYQHLIENSYSYVETRKATNKY-PTIANTONTQKGMYSQKELIS---- 1439
QY 448 GYLTPSPQFMD--PGLALHLAGTTRIGFDK-----ATTVADNNSL 486
Db 1440 -----KFLDCAVGLRNNLDEAQLSMQKKWENLKDSIAKTSAGNOCI 1481

RESULT 7
AMYG_ASPOP
ID AMYG_ASPOP STANDARD; PRT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GLAA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Euryotales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9125474; PubMed=1368680;
RA Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
RT "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
RT nucleotide sequence, and expression in Saccharomyces cerevisiae.";
RL Agric. Biol. Chem. 55:941-949(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92104497; PubMed=1761224;
RA Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
RA Hara S.;
RT "Nucleotide sequence and expression of the glucoamylase-encoding gene
RT (glaa) from Aspergillus oryzae.";
RL Gene 108:145-150(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=RIB 40;
RA Hara S., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
RA Hata Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D01035; BAA00841.1;

```

```

DR EMBL; D10598; BAA01540.1; -.
DR PIR; JQ1346; JQ1346.
DR HSSP; P04064; IGAI.
DR InterPro; IPR002044; CBD_4.
DR InterPro; IPR000165; Glyco_hydro_15.
DR Pfam; PF00686; CBD_4; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GLHYDRLASE15.
DR PROSITE; PS001568; CBD_4; 1.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT PROPEP 20 25 BY SIMILARITY.
FT CHAIN 26 612 GLUCOAMYLASE.
FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
FT DISULFID 236 239 INTERACT WITH SUBSTRATES (BY SIMILARITY).
FT DISULFID 248 475 BY SIMILARITY.
FT DISULFID 288 296 BY SIMILARITY.
FT CARBOHYD 39 39 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 612 AA; 65486 MW; CD7B23E5FA978F97 CRC64;

Query Match 3.3%; Score 97; DB 1; Length 612;
Best Local Similarity 20.5%; Pred No. 7.6;
Matches 105; Conservative 56; Mismatches 161; Indels 190; Gaps 28;

QY 78 ENLSAEAVTRGV---GGMSTHWTCSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELY 133
Db 107 EFTSSQARIQGISNPSGALSS-----GGLGEPKFNVDYTAFTGAMGRQP 150
QY 134 SEAERLIGTSTKEFDESI---RHTLVLSLQDAYKDRQIRFRLPLACRLHNAEYVFW 190
Db 151 RDGPALRATAMISFGELWLVENSHTSTATDL-----VMPV-----VRNDLSYVAQ 194
QY 191 HSAENLFHSIYNDKOKKLTLLTNHRCRLALTGG--YEKKIG-----AAEVRNL 239
Db 195 YNSQSGF-DLWEEVQGTSEFTVAVSHR---ALVEGSSFAKTVGSCPYCDSQAPQVRCY 249
QY 240 LATRNPSSOLDYIMAKVYVVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYITQ 299
Db 250 L-----QSFTGSGYIQAN-----FGGSRSGKDIN-----TVLGSIHTFD 283
QY 300 PMAFQOIVLRQEFVDSVRDDPYGLPMWKEAVAQHTAKNPTDALPIPF----- 346
Db 284 POATC-----DDATFQPC SARALANH--KVVTDSFRSYATNSGRAENQAVA 328
QY 347 --RDPEPQVTPPTTEHPWH-----TOIHRDAFSYGAVGPEVDSRVIVD--LRWF--- 392
Db 329 VGRYPEDS-----YNGNPNWFLTLAAAEQLYDALQWDKIG-----SLATDVSLEPFKAL 380
QY 393 ---GATDPEANLLVFQNDVQ-----DGYSMPQPTFRYPSTASNVNRKRMADMCEVA 443
Db 381 YSSAATGTVASSTTVVKDVSVAKVADGYVQIVQT--YAASTGS-----MAEQ---- 427
QY 444 SNLGYLPISPPQFMDPGLALHLAGTTRIGFDKATVADNNSLVDFANLYVAG---NGT 500
Db 428 -----YTKTDGSGO-----TSARD---LTWSYAALLTANNRRNAV 458
QY 501 IRTGFGEN-----PTLTSMCHAISARSINT 527
Db 459 VPAPWGETAATSIIPSNACSTTSAGTYSSVVI 490

RESULT 8
ID US02_HSV11 STANDARD; PRT; 291 AA.
AC P06485;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)

```

DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Protein US2.  
 GN US2.  
 OS Herpes simplex virus (type 1 / strain 17).  
 OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae: Simplexvirus.  
 OX NCBI\_TaxID=10299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=85160822; PubMed=2984429;  
 RX McGeoch D.J., Dolan A., Donald S., Rixon F.J.;  
 RA "Sequence determination and genetic content of the short unique  
 RT region in the genome of herpes simplex virus type 1.";  
 RL J. Mol. Biol. 181:1-13 (1985).  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES US2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L00036; AAA96686.1; -;  
 DR EMBL; X14112; CAA32279.1; -;  
 DR EMBL; X02138; CAA36056.1; -;  
 DR PIR; A05238; Q0BE72.  
 DR InterPro: IPR003485; US2.  
 DR Pfam: PF02476; US2; 1.  
 DR SEQUENCE 291 AA; 32470 MW; 7227D9F051FD3A78 CRC64;  
 SQ -----  
 Query Match 3.3%; Score 96; DB 1; Length 291;  
 Best Local Similarity 21.0%; Pred. No. 3.1;  
 Matches 73; Conservative 31; Mismatches 113; Indels 130; Gaps 17;  
 QY 236 VRNLLATRN--PSQLD-----SYIMAKYVVLASGAIGNPQILYNSGFSGLQ----- 280  
 Db 8 VMTLLDQNALPRTSDVSPALWSFLRLQCRILASEPLGTPVVRPANRLRLAEPLMDLP 67  
 QY 281 -----VTPRNDSLIPN--LGRYTEOPMAFCQIVLRQEFVDSV-----RDPD----- 320  
 Db 68 KPTRIVRTSCRCPPNTTGLFAEDSPLE-----STEVVDVACFRLRLHQPSPPRL 121  
 QY 321 YGLPWWKEAVA-----QHIAKNPTD-----ALPIPRDPPEQVTPP 356  
 Db 122 YHL--WVVGADLCVPFLEYAQIRLGVRTIAKTPDAWGEPAVTRFELPWTVAWTP 179  
 QY 357 F--TEEHPWHITQIHRDAFSYGAVGEVDSRVIVDLRFWFGATDPEANLLVFQNDVQDYS 414  
 Db 180 FPAAPNHPLETLRSRYEQYGVVLPGTNGRERDCMRWL-----RSLIALHK----- 225  
 QY 415 MPQTFYRSTASNVARKMADMCEVASNLGGLYFTSP---PQFMDPGALHLAHTTR 471  
 Db 226 -PHPA--TPGLTTSHPVRRPCCA--CMGMEPVDEQPTSEGRGQPTDPLIAVR----- 275  
 QY 472 IGFDKATTADNNSLVDFANLYVAGNGTITGTGNGPTLTSMCHAI 518  
 Db 276 -----GERPLPHICYPV 288  
 RESULT 9  
 ID YIN0\_YEAST STANDARD; PRT; 964 AA.  
 AC P40467;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative 108.8 kDa transcriptional regulatory protein in FKH1-STH1  
 DE intergenic region.  
 GN YIL130W.  
 OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Barrell B.G., Badcock K., Bankler A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
 RA Lewis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
 CC CLUSTER DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z38059; CAA86148.1; -;  
 DR PIR; S48404; S48404.  
 DR HSSP; P08657; 1CLD.  
 DR SGD; S0001392; YIL130W.  
 DR InterPro: IPR001138; Zn2\_Cy6\_fungal.  
 DR Pfam: PF001172; Zn\_c1us; 1.  
 DR PRINTS; PR00054; FUNGAL2NCYS.  
 DR SMART; SM00066; GAL4; 1.  
 DR PROSITE; PS00463; ZN2\_Cy6\_FUNGAL\_1; 1.  
 DR PROSITE; PS50048; ZN2\_Cy6\_FUNGAL\_2; 1.  
 KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Nuclear protein; Zinc; Metal-binding.  
 FT DNA\_BIND 21 47 ZN(2)-CYS(6), FUNGAL-TYPE.  
 FT DOMAIN 811 896 HIS-RICH.  
 SQ SEQUENCE 964 AA; 108780 MW; ADSABE59E4B022CC CRC64;  
 Query Match 3.3%; Score 96; DB 1; Length 964;  
 Best Local Similarity 19.5%; Pred. No. 18;  
 Matches 100; Conservative 76; Mismatches 163; Indels 174; Gaps 27;  
 QY 13 GYHKK-----NELEFQK-----DIDRFVNVKIGALQOVSPVRNQNVFTLD--- 53  
 Db 380 GFHRKLSPNSGFSPFIEIEMRKRLFTYTKLDVYINAMGLPRISPDQDTLP-LDLS 438  
 QY 54 -----PCAWSAPPGSSAISNGKNPHOREF-----ENLSAEAVTRGVGCMSPHWTCST 100  
 Db 439 ENITEVAYLPENQHSVLSSTGIS---NEHTKFLILNEIISLYPIKKTNSIISETWTS 495  
 QY 101 --PRIHPMESLPGIGPKLSN-DPAEDDKENNELYSEAEERLIGTSTKEFDESIRHTLV 157  
 Db 496 LELKRLNWLDSLKPLIPNAENIDPE-----YERANRLHL-----SPLHVQII 539  
 QY 158 RSLQDAYDKQRIFRPL-----PLACHRLKN-----APEYVE--- 189  
 Db 540 -----LYRPFHYLSRNMNAENYDPLCYRRARNISAVARTVIKAKEMVSNL 587  
 QY 190 -----WHSANLPHSIY-----NDDKOKKLTLLTNHRCRLALTGGYEKKIG 232  
 Db 588 LTGSYWTACTYTFYSVAGLLFYHEAQLPKDSAREYDILKDAETGRSVLIQKDSMA 647  
 QY 233 AAERYNLLATRNPSQLDSYIMAKYVVLASGAIGNPQILYN--SGFSGLOVTPRNDSLIP 290  
 Db 648 ASRTYNLL-----NQIFEKLSNKTQIITA-----LHSSPSNESAFLVNNSSALKP 693  
 QY 291 NLGRYTEOPMAFCQIVLRQEF-----VDSVRDDPYGLPWWKEAVAQIAKNPFDALPIP 345  
 Db 694 HLGDSL-QPPVFFSSQDTKNSFLAKSEESTND-----YAMANYLNNPTISENPLN 743

QY 346 FRDPERQV---TTPFTTEHPWHTQTHRDFAFSYGAVGPEVDSRVIVDLRWFATDPEANNL 402  
 Db 744 EAQQQDVSGQTTNSNE-----RDPNNFLSIDRLDNN-----GQSN 782  
 QY 403 LVFONDV---QDGYSMPQPT---FRYRPTASN 429  
 Db 783 LQATDDVFIRNDG-DIPTWNSAFDFSSKSNASN 814

RESULT 10  
 AMYH\_SACFI STANDARD; PRT; 519 AA.  
 AC P26989; P78745;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Glucoamylase GLAI precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).  
 GN GLAI.  
 OS Saccharomycopsis fibuligera (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.  
 OX NCBI\_TaxID=4944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ;  
 RX MEDLINE=92137640; PubMed=1840532;  
 RA Hostinova E., Balanova J., Gasperik J.;  
 RT "The nucleotide sequence of the glucoamylase gene GLAI from  
 Saccharomycopsis fibuligera KZ";  
 RL FEMS Microbiol. Lett. 67:103-108(1991).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=KZ;  
 RA Hostinova E.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES. ONLY DIFFERS IN 11 POSITIONS WITH GLUCOAMYLASE GLUI.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X58117; CAA41120.1; .  
 DR PIR; S14596; S14596.  
 DR HSSP; P08017; IAYX.  
 DR InterPro; IPR000165; Glyco\_hydro\_15.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
 DR PRINTS; PR00736; GLHYDLRLA15.  
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.  
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal.  
 FT SIGNAL. 1 27  
 FT CHAIN 28 519  
 FT BINDING 166 166  
 FT ACT\_SITE 234 234  
 FT ACT\_SITE 237 237  
 FT ACT\_SITE 238 238  
 FT CARBOHYD 115 115  
 FT CARBOHYD 127 127  
 FT CARBOHYD 205 205  
 SQ SEQUENCE 519 AA; 57542 MW; A15A009A7640053C CRC64;

Query Match

3.3%; Score 95.5; DB 1; Length 519;

Best Local Similarity 19.2%; Pred. No. 7.8;  
 Matches 107; Conservative 62; Mismatches 199; Indels 189; Gaps 25;

QY 20 IFQKIDIDFVNVIKALQOVSVPRVQNVPTLPDCAWAPGSSAISGNKP----- 72  
 Db 49 LDKQKDVSLYY-----LLQNIAYPEGQFNDGV--PQTVIASPST-----NPDIYYQMT 95  
 QY 73 -----HQREFENLSAEAVTRGVGGMSTHWTCTSPRIHPHMEPLGIGRPLK 118  
 Db 96 RDSAITFLVLSELDNNTTTLAKAVEYYINTSYNLQRTSNPSGSDFDENHKLGEPKF 155  
 QY 119 SNPDABDDKENNELYSEARL-----IGTSKKEFDESIRHTLVL-----RSLQDAYKD 166  
 Db 156 NTDGSAYTGAWRPQNDGPPALRAYAISRYLNDVNSLNKGLVLTDSGDINFSTEDIYKN 215  
 QY 167 RQIRFPLPLACHRLKNAPEYV--EWSAENLFHSTYNDKOKKLTLLTNHRCPLALT 224  
 Db 216 ---ILKP-----DLEYVIGWDSYG---FDLWEENQGRHFTSLVQOKA---LAYA 257  
 QY 225 GGYEKKIGAAEVRNLLATRNPSQSDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTPR 284  
 Db 258 VDIAKSFDGDFANTLS--STASTLESYL-----SGSDG-----GFVNTDV--- 296  
 QY 285 NDSLPLNLGRYITEQPMAPCQIVLRQEFVDSVRDDPYGLPWWKEAVAQHIANKPTDALPI 344  
 Db 297 -----NHIVENPDLLQ--NSRQGLDS-----ATYIG-----PL 323  
 QY 345 PFRDPEPQVTTPTTEHPWHTQTHRDFAFSYGAVGPEVDSRVIVDLRWFATDPEANNLV 404  
 Db 324 LPHDIGESSSTPF-----DVDNEYVL-----QSYLLLL 351  
 QY 405 FQN-----DVQDGYSPQPQTFRYRPTASTNVRARKMADMCEVASNLGGYLPTSPQFMDP 460  
 Db 352 EDNKORYSVNSAYSAGAAIGRY-PEDVYN-----GDSSSGNPFPLAT 393  
 QY 461 GLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGGENPTLSMCHAIKS 520  
 Db 394 AYAAQVPYKLVYDAKSASNDITINKINYDFNKYIVDLSTINSYG-----QS 440  
 QY 521 ARSIINTLKGDTGKNT 537  
 Db 441 SDSV--TIKSGSDEPNT 455

RESULT 11  
 HELS\_SULSO STANDARD; PRT; 708 AA.  
 ID HELS\_SULSO  
 AC Q97YI9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Putative sk12-type helicase (EC 3.6.1.-).  
 DE SS02462.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Dawez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,  
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.B., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;  
 RT The complete genome of the crenarchaeon Sulfolobus solfataricus P2.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: AE006844; AK42601.1; ALT\_INIT.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR003583; HHH\_1.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00270; DEAD; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; Helic; 1.  
 DR SMART: SM00278; HHH1; 1.  
 DR Hypothetical protein; Hydrolase; Helicase; ATP-binding;  
 KW Complete proteome.  
 FT NP\_BIND 46 53 ATP (POTENTIAL).  
 FT SITE 145 148 DEXH BOX.  
 FT SEQUENCE 708 AA; 80500 MW; D52C24C89541C780 CRC64;

Query Match 3.3%; Score 95.5; DB 1; Length 708;  
 Best Local Similarity 21.3%; Pred. No. 12;  
 Matches 53; Conservative 40; Mismatches 93; Indels 63; Gaps 11;

QY 69 GKPHQREFENLSAEAVTRGVGMSTHTCTSPRIH-----PPMESLPGIGRKLNSNDPAE 124  
 Db 492 GKRVDLYINPFTADIIRKLEGYKA--SCEIAYLHLAFTPDGPLYVSGR-----N 541  
 QY 125 DKEWNELYSEAE-RLTGTSKTFEDESIRHTLVLRSLQDAY-----KDRQIRFRPLP 175  
 Db 542 EEEELIELLEDCELVVERPYEEDS--YSLYNALKVALIMKDWIDEVDEDTILGKYN 598  
 QY 176 LACHRLKNAPEYVFW--HSAENLFHSTYNDKQKFLTLTNHRCRTRALTTGGYKKGIGA 233  
 Db 599 IGSGLDRIYVETDWTLYSAYHLSKELRLDDHSDKL-----RLNLRVTDGVEEL-- 649  
 QY 234 AEVRNLLATRNPSQLDSYIMAKVYVILASGAIG--NPQILYNSGFGSLQVTPRNDSLIPN 291  
 Db 650 -----LELVQIGVGKRRARLLYNGIKGLGDVWVNPDRVN 686  
 QY 292 -LGRYTEQ 299  
 Db 687 LLGOKLGER 695

RESULT 12  
 ID CCAA\_BACTU STANDARD; PRT; 1257 AA.  
 AC Q45754;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Pesticidal crystal protein cry12Aa (Insecticidal delta-endotoxin  
 DE CryIIa(a)) (Crystalline entomotoxin)  
 DE protein.  
 GN CRY12AA OR CRYXIIA(A) OR CRYVB.  
 OS Bacillus thuringiensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1428;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.  
 RC STRAIN=NRRL B-18244 / PS33F2;  
 RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,  
 RA Sick A.J.;  
 RT "Novel Bacillus thuringiensis microbes active against nematodes, and  
 RT genes encoding novel nematode-active toxins cloned from Bacillus  
 RT thuringi.";  
 RL Patent number EP0462721, 27-DEC-1991.  
 CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.  
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTALLINE PROTEIN IS PRODUCED DURING

CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
 CC OF THE SPORE COAT.  
 CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
 CC N-TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: L07027; AAA22355.1;  
 DR InterPro: IPR001178; Endotoxin.  
 DR Pfam: PF00555; endotoxin; 2.  
 KW Toxin; Sporulation.  
 SQ SEQUENCE 1257 AA; 142265 MW; 3D988BFFC6C0E3981 CRC64;

Query Match 3.3%; Score 95.5; DB 1; Length 1257;  
 Best Local Similarity 20.7%; Pred. No. 29;  
 Matches 120; Conservative 78; Mismatches 206; Indels 175; Gaps 34;

QY 12 PGYHKNEIEFKDIDRFVNVIKALQOVSVVRNQNVPT-LDPGAWSPGSSAISNGK 70  
 Db 359 PNTHETVHVFSTYNTQSSGNISRGSNPIPIDLNPIISTCIRNSFYKATGSSVLNFK 418  
 QY 71 NPQOREFENLSAEAVTRGVGMSTHTCTSPRIHPPHESLPGIGRKLNSNDPAEDKEWN 130  
 Db 419 DGTQ---GYAFQAQPTGGA-----WD-----HSFTES-----DGAPEGHKLN 452  
 QY 131 ELYSEAEERLTGTSKTFEDESIRHTLV-LRSLQDAYKQRIERPLPLACHRLKN----- 183  
 Db 453 YITSP---GDLRDFINV--YTLISTPTINELSTKIKGF---PAEKGYIKNQINKY 503  
 QY 184 --APEYVHSAENLFHSIYNDKQKFLTLTNHRCRTRALTTGGYKKGIGAEEVRL-- 239  
 Db 504 YGKPEYINGAQPVN-----ENQOTLIFE-----HASKTAQYIRIRYASTQGTGK 550  
 QY 240 -----LATRNPSQLDSYIMAKVYVILASGAIGNPQILYNSGFGSLQVTPRNDSLIPN 291  
 Db 551 YRLDNOELQTLNPTSHNGYV-----TGNIGENYDLYTIG--SYTITEGNHTL--- 597  
 QY 292 LGRYTEQPMACQI-----VL-ROEFV--DSVRDDPYGLPWKEAQAHIKAKNPTD 340  
 Db 598 -----QIQHNDKNGVLDRIEFVPRDSLQDSP----- 624  
 QY 341 ALPIPRDPEPQV---TTPFTEHP--W-----HTQIHRDAFSYGAVPEVDSRVIVDL 389  
 Db 625 -----QDSPPEVHESTIIFDKSSPTIWSNKHSHYSHIHLEG-SYTSQG-SYPHNLINL 676  
 QY 390 RWFQATDPEANLLVFN--DVQDGYSMQPQTFRYPSTASVNRARKMADCEVASNLG- 447  
 Db 677 --FHPTDNRNHTIHVNGDMNVYD-----KDSVADGLNFKNTATIPSDAWSGT 726  
 QY 448 -----GYLTPSPQFMDPGDLALHLAG--TTRIGFDKAT---TVADNNSLVDFANL 493  
 Db 727 ITSMHLENDNNFKITPKF---ELSNELNITITQVNFALFASSAQDTLASVSDYV--TEQV 782  
 QY 494 YVAGNGTIRTFGG--ENPTLTSMCHAISARSIIINTLKGK 531  
 Db 783 VMKVDALSDVEFGKEKKALKRLVNAQAKRLSKIRNLLIGG 821

RESULT 13  
 ID YDAK\_YEAST STANDARD; PRT; 500 AA.  
 AC P28817;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

```

RESULT 14
CDH_PHACH
ID CDH_PHACH STANDARD; PRT; 773 AA.
AC Q01738; O00047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cellulose dehydrogenase precursor (EC 1.1.5.1) (CDH) (cellobiose-
DE quinnone oxidoreductase).
DE CDH-1 AND CDH-2.
GN Phanerochaete chrysosporium.
OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Corticiaceae; Phanerochaete.
OX NCBI_Taxid=5306;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=OGC101;
RC MEDLINE=97077226; PubMed=8919793;
RA Li B., Nagalla S.R., Renganathan V.;
RX "Cloning of a cDNA encoding cellobiose dehydrogenase, a
RT hemoflavoenzyme from Phanerochaete chrysosporium.";
RL Appl. Environ. Microbiol. 62:1329-1335(1996).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=OGC101;
RC MEDLINE=97176414; PubMed=9023960;
RA Li B., Nagalla S.R., Renganathan V.;
RX "Cellobiose dehydrogenase from Phanerochaete chrysosporium is encoded
RT by two allelic variants.";
RL Appl. Environ. Microbiol. 63:796-799(1997).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 19-208.
RP MEDLINE=20139694; PubMed=10673428;
RX Hallberg B.M., Bergfors T., Boeckbro K., Pettersson G., Henriksson G.,
RA Divne C.;
RX "A new scaffold for binding haem in the cytochrome domain of the
RT extracellular flavocytochrome cellobiose dehydrogenase.";
RL Structure 8:79-88(2000).
CC -1- FUNCTION: DEGRADATES BOTH LIGNIN AND CELLULOSE. OXIDIZES CELLOBIOSE
CC TO CELLOBIONOLACTONE.
CC -1- CATALYTIC ACTIVITY: Cellobiose + a quinone = cellobiono-1,5-
CC lactone + a phenol.
CC -1- COFACTOR: ONE FAD AND ONE HEME B.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMC
CC OXIDOREDUCTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U46081; AAC49277.1; -
DR EMBL; U65888; AAB61455.1; -
DR EMBL; U50409; AAB92262.1; -
DR PDB; 1D7B; 18-OCT-99.
DR PDB; 1D7C; 18-OCT-99.
DR PDB; 1D7D; 18-OCT-99.
DR InterPro; IPR000172; GMC_oxred.
DR InterPro; IPR001100; pyr_redox.
DR Pfam; PF00732; GMC_oxred; 1.
DR PRINTS; PR00411; PNDROTASEI.
DR PROSITE; PS00623; GMC_OXRED.1; 1.
DR PROSITE; PS00624; GMC_OXRED.2; 1.
KW Cellulose degradation; Oxidoreductase; FAD; Flavoprotein; Heme;
KW Multigene family; Signal; 3D-structure.
FT SIGNAL 1 18
FT CHAIN 19 773 CELLOBIOSE DEHYDROGENASE.
FT FT

```

	-!	FUNCTION:	TRANSCRIPTIONAL ACTIVATOR.
CC	-!	SUBCELLULAR LOCATION:	Nuclear (Probable).
CC	-!	SIMILARITY:	CONTAINS 1 GATA-TYPE ZINC FINGER.
CC	--	-----	-----
CC	This	SWISS-PROT entry is copyright.	It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licensee@isb-sib.ch).
CC	--	-----	-----
DR	EMBL;	AL035076;	CAB22647.1; .
DR	EMBL;	AL049521;	CAB40003.1; .
DR	EMBL;	L31601;	AAC35593.1; .
DR	HSPS;	P17429;	AGAT.
DR	TRANSFAC;	T02831;	.
DR	InterPro;	IPR000679;	ZnF_GATA.
DR	Pfam;	PF003320;	GATA; 1.
DR	PRINTS;	PR00619;	GATAZFINGER.
DR	SMART;	SMO0401;	ZnF_GATA; 1.
DR	PROSITE;	PS00344;	GATA_ZN_FINGER_1; 1.
DR	PROSITE;	PS00114;	GATA_ZN_FINGER_2; 1.
KW	Transcription regulation; Activator; DNA-binding; Zinc-finger; Nuclear protein.		
FT	ZN-FING	635	GATA-TYPE.
SEQUENCE	855 AA;	91776 MW;	3D932F83D2DE677A CRC64;
<hr/>			
Query Match            3.3%; Score 94.5; DB 1; Length 855;			
Best Local Similarity     20.7%; Pred.No. 20;			
Matches 121; Conservative       59; Mismatches 197; Indels 207; Gaps      31;			
<hr/>			
Qy	64	SAISGNKNPHOREPE----	NLSAEAVTRGVCGHSTHTWCSTPRIHPMESLPGLGPKLS 119   :        :    :
Dd	50	SELNGRRVENVLRMLMSINLKNTPGDSNTLTPTDFSDTPAPSSAOVSVPPTSAAET 109   :        :    :	
Qy	120	NDPAEDDKENEL--YSEALERLICTSKE-----	FDSEIRHTVLRLSLADA-Y 164   :    :    :    :    :
Dd	110	ADNSDTMKLNPIPAYSVPADTTGSLLMEFYIQRVRKRTSFDESTAKS-KRKSIADSHE 168   :    :    :    :    :	
Qy	165	KDRIIFRPPLACHRLKNAP-EYVEWHSAEHLFSHYINDKKOKKLFTLLTNHRCTRALL 223   :    :    :    :    :	
Dd	169	PDPNAMORP-----HDLESQPFPFKIH-ASNSFNFKVD-	----- 202   :    :    :    :    :
Qy	224	TGGYEKGIGRAEVNNLATRNPPSQOLDYSITMAKVYLASCIGNPOLLYNSGESGLQTVP 283   :    :    :    :    :	
Dd	203	-----IDSNFNLDASALPIPPSPDDFFSHNLNPAN---PSPANS-----	243   :    :    :    :    :
Qy	284	RNDSLIPNLGRYITEQQPMACQIVLRQEFVDSRDDPY-----GLPWKEAVAQHIAKN 337   :    :    :    :    :	
Dd	244	NNSASN-N-QRIKASKPHADTVLGIDLF-DMTPESEPSPENGFGPSFDVAANTHQTLTF 299   :    :    :    :    :	
Qy	338	PTDA-----LPPIFRPEPVQVTTFTEEHFWHTQIHRDAFS-----YGAV 377   :    :    :    :    :	
Dd	300	PSSATNSFESEHGSGAPPPIP-----GSPVPSIYA-H NTASEDGESSSYNQGLEGIS 350   :    :    :    :    :	
Qy	378	GPEVDSRVVIDLRFWGATDEANNLL-VFOOD-----VQD-----GY 413   :    :    :    :    :	
Dd	351	SP-LSSGTVPNQSFF--PDVSGNNIFDVSRNRNHVESPLIQSPGSYVMPSINMVSSLPI 407   :    :    :    :    :	
Qy	414	SMPOP-----TFR-----YRPSTATNVARKMMADM 439   :    :    :    :    :	
Dd	408	SAPVPSNOSPFRPNPTRFRNSSKVGCSGGSDVNNOENAESTNPSSIHSINSLEWASGET 467   :    :    :    :    :	
Qy	440	CEVASNLGGVIPT-S--PPQMDFPGLAHLAGTTRICGFDAKTATVADONNISLVDMFLNYVA 496   :    :    :    :    :	
Dd	468	TGHSSN-SPLPGSDMESPQFMRVGTAMGA-----PYRSNSS-----	503   :    :    :    :    :
Qy	497	NGTIIRTGFGEN-----PILTSMCHAIKSARSIIINTLKGTDG 534   :    :    :    :    :	
Dd	504	-----NFNGQNFPHTSPQFSAVPHRKVSQAQD--TNLMGSSPG 539   :    :    :    :    :	

Search completed: September 3, 2002, 16:01:29  
Job time: 338 sec







Db 506 MVGMRKITEANPMKIIG--AKPGTHFDTTVQTHMSGGAIMGEDPKTSVNRYLQ 563  
QY 487 VMDPANLYVAGNGTIRGFGSNPT 510  
Db 564 SMDVNVFVPGASAFPQGLGYNPT 587  
RESULT 4  
A84260  
hypothetical protein Vng1035c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: A84260  
R:Mg, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabcig, Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: A84260  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-529 <STO>  
A:Cross-references: GB:AE004437; MID:g10580587; PIDN:AGI9445.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG1035C

Query Match 5.5%; Score 158.5; DB 2; Length 529;  
Best Local Similarity 22.3%; Pred. No. 0.00036;  
Matches 117; Conservative 44; Mismatches 194; Indels 169; Gaps 22;  
QY 84 AVTRGVGMSTHTWCSTPRIHP---MESLPGRKPLSDNPAEDDKWENLYSEARLI 140  
Db 88 ARVKGVGSTLHWGMVRLHEQDFRLASATGVA---DWPFDYDTLRPYAAASAL 142  
QY 141 ---GTSKEFDESRLVLSLQDAYKDRQIRPRPLACHRLKNAPEYVEHSAENLF 197  
Db 143 GVSASDNFPAPPEQPHQPAFPSPSYSD--SLFAD---ACESIGIAT----- 185  
QY 198 HSIYNDKQKKLFTLLNHCRTRALG---GYE-----KKIG 232  
Db 186 HSPVN-----ARLSAGRETRACVGYGTCQPCVPSGAKYATVHVDRATDAGARVID 237  
QY 233 AAVERNL-----LATRNPSSQLDSYIMARVYVLASGAINPOL-----Y 272  
Db 238 EAPVORLEHDAAGDRVTGAVYATPDGTHROS---ATEFVLAAGGIETPRLLLLSDSDRY 294  
QY 273 NSGF---SGLOVTPRNDLIPNLGRYTEOPMAFCQIVLRQEFVDSVRDDPYGLPWNKEA 329  
Db 295 PDGLANSSGL-----VGRY-----FMD----- 311  
QY 330 VAQHIANKPTDALPIPRDPEPQVTTFTTEHPHHTQIHRDAFSYGAV-----GPEV 381  
Db 312 ---HLFAGAGTLDPETRQNHVGENT---TESHQY---DRPDGSRGAIKLEFLNYAGPSP 363  
QY 382 DSRVIVLRWFGA-----TDPEANLLVFQNDVQCYSPQPTFR 421  
Db 364 AEMALSGDDGDMACDRIDASGTHIAGVGLVEQOPRENVRVLRHPERTDVHGNFVDDV 423  
QY 422 YRPSTASNVARKMMADCEVASWLGGLP--TSPQPFMDPGLALHLAGTTRIGFDKATTV 480  
Db 424 WLSAYERRTIERANEIQREILLTELGAIEWTVGPE--DTGPAPHMGHTTRMGTDPAESV 481  
QY 481 ADNNLSVDFANLYVAGNGTIRGFGSNPTLTSCHAIKARSII 524  
Db 482 VDPRLRTHLSNLSVASSSVFPTAGAMNPTLTAAALKAADHI 525

RESULT 5  
AB2003  
dehydrogenase chain [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AB2003  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigui, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2003  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-619 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA077942.1; PID:g17135396; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr1576

Query Match 5.5%; Score 158.5; DB 2; Length 619;  
Best Local Similarity 19.7%; Pred. No. 0.00046;  
Matches 125; Conservative 72; Mismatches 217; Indels 221; Gaps 27;  
QY 22 FQKDIDRFVNVIKALQOVSVPRNQNVPTLDGAMSAAPPGSSA-----ISNGKN 71  
Db 60 FQSYVDTFYKAVD-----KNPNSP-----YPANSNVQSPDNDYFIEQGP 101  
QY 72 PHOREFENLSAEAVTRGVGMSTHTWCSTPRIHP---PMESLPGRKPLSDNPAEDDKKE 128  
Db 102 P-----LAGSYTRVLGGTTHWEAKTPRMLPDEPKLSSTYGGGL-----DWPIDYHD 148  
QY 129 WNELYSEARLIGTSTK-----EFDESIRHTLVLSLQDAYKDRQIR----- 170  
Db 149 LEPYRKAHEHMGVCGVDQORALGLEFPQD--VYFPMKLPSPSYLDQKVKIEKNGTVNE 206  
QY 171 -----FRPLACHRLKNAPEYVEHSAENLF-----HSIYNDK----- 205  
Db 207 LYGKTHLSFSTFPQARNGVEN-PKY---DQGNLFVPDGTVTSVHPVQYGERCOGNANCY 261  
QY 206 -----QKKLFTLLNHCRTRALG---GYE-----KKIG 250  
Db 262 PICVPQAKYDARRTLKSAFETGKVHVLYQAVAYKVEYDRQTRITAIHKYKKPNSSEY 321  
QY 251 SYIMAK--VYVLASGAINPQILYNSGFSGLQVTPRNDLIPNLGRYTEOPMAFCQIVL 308  
Db 322 TTGIAKGTFLVATNAVENARLLIGSDL-----PNTSRLI---GRYL----- 360  
QY 309 RQEFVDSVRDDPYGLPWNKEAVAQHIANKPTDALPIPRDPE-----POVTTTF----- 357  
Db 361 -----MDHPFTLAW-----ALM-----PEVTGTRGRLVTSIGITFR 392  
QY 358 -----TEHPHHTQIHRDAFSYGAVGPEVDSRVIVD-----LR 390  
Db 393 KGDFRKKQSAFANDIHDNGMGWATGSPKSEVEDADVANKNKYQBELRQTLISRIERQLLA 452  
QY 391 WFGATDPEANLLVFQNDVQCYSPQPTFRYRPSTASNV--RARKMMADCEVA-----S 444  
Db 453 FMCELLPEYGNRVITIDPRHDKLG-----NYPVINFNLDPYSRRLAYTRKYSRVIFE 506  
QY 445 NLGG-----YLPSPQPFMDPGLAL-----HLAGTTRIGFDKATTVADNNLSVDFANL 493  
Db 507 RLGAEDYTHDPQDPAYFEFEGEYVYKGGNHFSGTHIMGTPLNSVVDYSLRSWDHKNL 566  
QY 494 YVAGNGTIRGFGNGNPTLTSCHAIKARSIIINTL 528  
Db 567 FLVGAGSMPTIGSSNTTLTAAALSFRTAEHMLQEL 601

RESULT 6  
B75608  
GMC oxidoreductase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: B75608  
R:White, O.; Eelsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: B75608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-722 <WHI>  
A:Cross-references: GB:AE001862; GB:AE001825; NID:96460468; PIDN:AAF12230.1; PID:9646052  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0127  
A:Map position: 2

Query Match 5.0%; Score 144; DB 2; Length 722;  
Best Local Similarity 21.1%; Pred. No. 0.0077;  
Matches 107; Conservative 68; Mismatches 196; Indels 136; Gaps 26;  
QY 90 GCMSTHTCSTPRIHPMESLPGICRPLSN-----DPAEDDKENNELYSEARLIG 141  
DB 277 GGSTVWMSV-----PPRDDI-----QRWASEHGLSDVADPGYD-----RHIDAVLERM-G 323  
QY 142 TSKTKFDESIRHTLVLSLQDAYKQRIFRPLPLACHRLKNAPEYVEHSAENLFHSYI 201  
DB 324 VSEQSDHNGPHQ---RLVEGADKLGYFVKA-----ALNLSPEH---YDADKAGHAGF 371  
QY 202 NDD---KQKLLFT-----LLTNHRCRRLATGGYKKGKIGAAAEVNRNLLATRNPS 247  
DB 372 GDGTGAKQGTLTFLKDAPEAGARILVGTQAQRLVLEDG-----RAAGYSAVTTMGDETR 426  
QY 248 OLDSYIMAKVYVLSAIGNPOLYNSGSLQVTPRNDLSLPNLRGYITEOPMAFCQIV 307  
DB 427 QIT---VRAPQVVVACAEPTALLRSIGG-----PAAGRYLRHLPAGLVAGI 473  
QY 308 LRQEFVDSVRDDPYG---LPWKEAQAQIAKNPTDALPIPRDPPE-----P 351  
DB 474 -----YGEDQRAWGP-----PQSGILKQFADHENGHGFIEGVQYGP 511  
QY 352 QVTT---PPT---EEH-PWHTQTHRDAFSYGAVGPEVDSRVIVD-----LRWFGATDP-BA 399  
DB 512 ALMASGLEPWTGGEAHRDLMSKFRHATFVSIQDRGHGQVTVDDGNAVHTYALTDLLDA 571  
QY 400 NLLYFONDVDGYSM-----PQPTFRYRPSTASNRARKMADMCVAS-----NLGGYLP 451  
DB 572 RN---FRRGVTSIRLHERAGAEIIVALAPGVAPWRGDDLEAFIGQVAQVPLGAGGQTV 628  
QY 452 TSPQFMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGGENPTL 511  
DB 629 FS-----AHQMSARMGSDPQTSVADPDGGLHDVPGVWIGDTSAPFTCSGVNPMV 678  
QY 512 TSMCHAIKARSIIINTLGGTDCKNTG 538  
DB 679 SCMALASRTAEKLAAMEGADGGTGG 705

RESULT 7  
H83362  
glucuronate dehydrogenase PA2265 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: H83362  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: H83362  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-591 <STO>  
A:Cross-references: GB:AE004652; GB:AE004091; NID:g9948287; PIDN:AAG05653.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2265

Query Match 4.9%; Score 141; DB 2; Length 591;  
Best Local Similarity 19.9%; Pred. No. 0.0096;  
Matches 120; Conservative 63; Mismatches 235; Indels 184; Gaps 22;  
QY 45 RNQNYPTLDPGAW-----SAPPGSSAISNGKNPHQREFENLSAEAVTR 87  
DB 38 RGENRDTYPDGAYPNTLDLTNTYTRGKLFQNLKSTVTSIRHGINDALTALPYRQLSAFLPCD 97  
QY 88 GVGGMSTHTCSTPRIHPMESLPG-----IGRPKLSNDPAEDDKENNELYSEAE 137  
DB 98 GYGGAGLHWSGVHFRIMPPEELRSHRYEERYGKFIPEGMTIQDYGVSYEELEPHFDFAE 157  
QY 138 RLIGTSTKEFDESIRHTL-----VLRSLQDAYKQRIFR----- 172  
DB 158 KYFGTS-----GTAHTYKGVVGVKGNFPAADRSDDFPLPALRQVYS--AQLFKAAEEL 209  
QY 173 -----PLPLA-----CHRLKNAPEYVEHSAENLFHSIYNDKOKK 208  
DB 210 GLHPYDLPAANASGPWNPYGVQMGPCNFCGFCSGVACVYMYSKASPNL--NILPALRQTP 267  
QY 209 LETILTNRCTRRLATGGYEKKIGA---AEVRNLLATRNPSQLDSYIMAKVYVLSGA 264  
DB 268 LPELRANCNVLKNLSDSGRQATGYTVYDAQGREIVQP-----AKLVIISAFQ 315  
QY 265 IGNPOLLVNSGSLQVTPRNDLSLPNLRGYITEOPMAFCQIVLRQEFVDSVRDDPY--- 321  
DB 316 FHNVRLLLSGI-GKPYDPTGEGV--VGKNFAYQNM-----TIKAFDQKVHTPFFVGT 368  
QY 322 -----GLPMWKEAQAQIAKNPTDALPIPRDPPEQVTPPF 357  
DB 369 GGGGVAVDDFNADHDGFLGVGGSPMW-----VQAGSKPIGGLAVPPGTS----- 417  
QY 338 TEEHPWHT---QIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPPEANNLLVFONDVDGYS 414  
DB 418 -----WGSQWQAVKDAYTH-----TVSDAHGSGNNMTYRDNYLDLDPTYKDAYG 461  
QY 415 MP--OPTERY-----RPSTASNRARKMADMCVA-----SNLGGYLPSPQFMDPCL 462  
DB 462 QLLRMTFDKWDKNEIRMSRYVTEHRKKAEMNPKAISVSKNFGDHENTRYQ----- 515  
QY 463 ALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGGENPTLTSCHAIKARS 522  
DB 516 TTHLLGGAIMGSDPKTSVLNRYLQSWDVHNVFVGASAFPQGTGYNPTGLVAALAYWSAK 575  
QY 523 SI 524  
DB 576 AI 577

RESULT 8  
JC7628  
glucoside 3-dehydrogenase (EC 1.1.99.13) - Halomonas sp. alpha-15  
C:Species: Halomonas sp. alpha-15  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
C:Accession: JC7628; PC7123  
R:Kojima, K.; Tsugawa, W.; Sode, K.  
Biochem. Biophys. Res. Commun. 282, 21-27, 2001  
A:Title: Cloning and expression of glucose 3-dehydrogenase from Halomonas sp. alpha-1  
A:Reference number: JC7628; MUID:21164693; PMID:11263965  
A:Accession: JC7628  
A:Molecule type: DNA  
A:Residues: 1-562 <KOJ>  
A:Accession: PC7123  
A:Molecule type: protein  
A:Residues: 2-11 <KO2>

C;Comment: This enzyme, as an oligomeric enzyme composed of catalytic and electron transfer and disaccharides.

C;Genetics:

A;Gene: g3dh

C;Keywords: oxidoreductase

Query Match 4.8%; Score 140.5; DB 2; Length 562;  
Best Local Similarity 23.7%; Pred. No. 0.0097;  
Matches 45; Conservative 29; Mismatches 77; Indels 39; Gaps 5;

Qy 372 FSY-GAVGPEVDSRVIVDLR-----W-----FGATDPEANLLVFQNDVQ 410

Db 379 FGOGAASREGWDREIAELNIGADLKALQPGGWITGMTGFGEMLPDHDNRISLDHSVR 438

Qy 411 DGYSMPQPTFRYRSTASNVARKKMDMCE-----VASNLGGYLPTSPQFMDPG 461

Db 439 DKWGLPVLISIDVELKQNERDMRDVQDAVDLLEAGKVKNGDVGDA-----PG 489

Qy 462 LALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSMCHAIKSA 521

Db 490 MGIHEMCTARMGRDPKTSVLNSHNQVWDAPNVFTDGCACMTSSSCVNPSTYMTALTARAV 549

Qy 522 RSIINTLKGG 531

Db 550 DYAVEELKRG 559

RESULT 9

E98192

Probable oxidoreductase chain CJ0415 [imported] - Agrobacterium tumefaciens (strain C58,

C;Species: Agrobacterium tumefaciens

C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002

C;Accession: E98192

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 234, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A;Reference number: A97359; PMID:11743194

A;Accession: E98192

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-561 <KUR>

A;Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170

C;Genetics:

A;Gene: AGR\_L\_980

A;Map position: linear chromosome

Query Match 4.7%; Score 135; DB 2; Length 561;

Best Local Similarity 24.8%; Pred. No. 0.026;

Matches 39; Conservative 25; Mismatches 59; Indels 34; Gaps 3;

Qy 392 FGATDPEANLLVFQNDVQDGYSPQPTFRYRSTASNVARKKMDMCE----- 441

Db 420 FGEMLPYHDNRVKLDHDKDKWGL-----PVLNMVEMKQNELDMRDVNDAYEMF 471

Qy 442 -----VASNLGGYLPTSPQFMDPGIALHLAGTTRIGFDKATTVADNNSLVWDFANLY 494

Db 472 EAVGIKNVKPSRGTYA-----PCMGHEMCTARMGRDPKTSVLNGNQVWDAPNVF 522

Qy 495 VAGNGTIRTFGENPTLTSMCHAIKSARSINTLKGG 531

Db 523 VTDGCACMTSACVNPSTYMTALTARAAEFAVSRKKG 559

RESULT 10

AE3094

oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C;Accession: AE3094

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, J.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl  
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; PMID:11743193

A;Accession: AE3094

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-561 <KUR>

A;Cross-references: GB:AE008689; PIDN:AAL45171.1; PID:g17742847; GSPDB:GN00187

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atu4377

A;Map position: linear chromosome

Query Match 4.7%; Score 135; DB 2; Length 561;

Best Local Similarity 24.8%; Pred. No. 0.026;

Matches 39; Conservative 25; Mismatches 59; Indels 34; Gaps 3;

Qy 392 FGATDPEANLLVFQNDVQDGYSPQPTFRYRSTASNVARKKMDMCE----- 441

Db 420 FGEMLPYHDNRVKLDHDKDKWGL-----PVLNMVEMKQNELDMRDVNDAYEMF 471

Qy 442 -----VASNLGGYLPTSPQFMDPGIALHLAGTTRIGFDKATTVADNNSLVWDFANLY 494

Db 472 EAVGIKNVKPSRGTYA-----PCMGHEMCTARMGRDPKTSVLNGNQVWDAPNVF 522

Qy 495 VAGNGTIRTFGENPTLTSMCHAIKSARSINTLKGG 531

Db 523 VTDGCACMTSACVNPSTYMTALTARAAEFAVSRKKG 559

RESULT 11

T10651

hypothetical protein T5F17.20 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Sep-2000

C;Accession: T10651

R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban

submitted to the Protein Sequence Database, June 1999

A;Reference number: Z16533

A;Accession: T10651

A;Molecule type: DNA

A;Residues: 1-748 <BEV>

A;Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20

A;Experimental source: cultivar Columbia; BAC clone T5F17

C;Genetics:

A;Gene: ATSP:T5F17.20

A;Map position: 4

A;Introns: 70/3

C;Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160

Query Match 4.4%; Score 128.5; DB 2; Length 748;

Best Local Similarity 19.7%; Pred. No. 0.13;

Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;

Qy 76 EFENLSAEAVTRGVGMSTHWCTSPRIHPHMESLPGIGRPKLSNDPAEDDKENNELYSE 135

Db 301 KFMLLAGSAVG---GCTAVNWSAS---IRTPDHVL-----QEW-----SE 334

Qy 136 AERLIGTSTKEF---DE-SIRHTLVLSLQDAYKDRQRIFRPLPLACHRL-----K 182

Db 335 GSKIKFFGSQEQSAMDEVITRIGTCRVKHGFN-----QVLRKGCERLGLAVESVPR 389

Qy 183 NAPE-----YVEWHSANLPHSIYDDKKQKLLFTLLTNHRCRLAL----TGGYEKK 230

Db 390 NSPDHYCGCGYCGACAGAKNGTDQTLWLDVAVENGAVILTGIAERFVLVDNNTSSNERK 449

Qy 231 -----IGAAEYRNLLATRNPFSSQLDSYIMARVYVYVLAIGAIGNPQILYNSGFGSLQVTPRN 285

```
Db 450 KRCVGVFASVGGKIGK-----FIEARVTVSAGSLTTPPLMSGLKN-----495
QY 286 DSLPLNLR-----YTEOPMAFCQIVLRQEFVDSVR--DDPYGLPWNKEAVA 331
Db 496 ----PNIGRNKLHPVLTWGFPEKDSFSGKNYEGGIITSVHHND-----TESCC 544
QY 332 QHTAKNPTDALPTPFRDPEPQVTPPTTEBHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRW 391
Db 545 KAIENPLIG-PASVAGLSPWVS-----GPDKERMIR-----576
QY 392 FGATDEANLLVFQNDVQGYSM--PQTFYRSTASNVARKMMADMCEVAS---NL 446
Db 577 YGRT-----AHLFALVRDLGSGVMENEVYTRTKDRENLAGLRQALRVSAAGAVEV 632
QY 447 GGY-----LPTSPPOFMDPGLAL-----HLAGTTRIGFDKAT 478
Db 633 GYRSDGQRMKCEAITKEAMEFELDEVDVAGVGKGEYWTYFSAHQMSGRMGVTAAEE 692
QY 479 TVADNNSLVMDFANLYVAGNGTIRGTGNGNPTLTSMCHAKSARSINTLKGT 532
Db 693 GALLDENGESWEAEGFLVCDGSLPSAVGVNPMITIQSTAYCISKIVDSIQNK 746

RESULT 12
dehydrogenase chain I (AF068066) [imported] - Agrobacterium tumefaciens (strain C58, Cer
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: E98202
R:Goodner, B.; Hinkie, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: E98202
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89143.1; PID:g15158953; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1138
A:Map position: linear chromosome

Query Match 4.2%; Score 122; DB 2; Length 527;
Best Local Similarity 19.8%; Pred. No. 0.23;
Matches 101; Conservative 55; Mismatches 175; Indels 180; Gaps 22;

QY 75 REFENLSABAVTRGVGMSTHWCSTPRHP---PMESLPGIGRPKLSNDPAEDDKWNE 131
Db 104 RDFREVEHE-----DGVSPAMPVSYAELEPPYAEARLFGV-RGRAGDDPTEPPRSAPY 156
QY 132 LYSEA--ERLIGTSKEPDESIRHTLVLSLQDAYKDRQIRFRPLA-----CHRL 181
Db 157 MHAPIHEPVGIRVAKGFER-----LGLRPFH-----MPSAIDYGPGLCRRC 199
QY 182 KNAPEYVWHSANLFSHSIYNDKOKKLF-----TLTNHRCRLALGGYKKIG 232
Db 200 GTCDAFVCFDAKG-----DAETRLRLPALRHPNVSLLTGARVRL--IADGGKHIV 250
QY 233 AAEVRLNLLATRNPPSOLDOSYIMAKYVVLASGAIGNPQILYNSGSLQVTPRNDSLIPN- 291
Db 251 AVEIER-----AGEITT-IEAPLFLVSAGAINSALLIRSA-----DEKKPNG 292
QY 292 -----LGRYTEQPM-----FCQIVLRQEFVDSVRDPP-----320
Db 293 LANSVGVGRYLMNHLGSLMGLLPFTINDTRFPKTMNSLNDFFDGTGDEAARGNVQMLG 352
QY 321 -----YGLFWKEAVAQHIKAKNPTDALPIFRDPEPQVTPPTTEHPWHTQIHRDAF 372
Db 353 NIQGPIMIRAAVPMWRPLANLLARHSVDFLVM-----SED-----387
QY 373 SYGAVGPEVDSRVIVDLRFWFGATDPPEANLLVFQNDVQGYSMPOPTFYRSTASN--- 429
Db 388 ----TPKYDSRV---KPGWKNAG-----LIYRPGDREAHQR 417
QY 430 -VRARKMMADMCEVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHMRSLLRKNGFPVVLGHSGFIEAPS-----HOCGTVMRGDDPKPEAALNALCOTY 469
```

```
QY 373 SYGAVGPEVDSRVIVDLRFWFGATDPPEANLLVFQNDVQGYSMPOPTFYRSTASN--- 429
Db 388 ----TPKYDSRV---KPGWKNAG-----LIYRPGDREAHQR 417
QY 430 -VRARKMMADMCEVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHMRSLLRKNGFPVVLGHSGFIEAPS-----HOCGTVMRGDDPKPEAALNALCOTY 469
QY 489 DFANLYVAGNGTIRGTGNGNPTLTSMCHAK 519
Db 470 DHPNLYVDAGFFPSSAALNPALTVAQAIR 500

RESULT 13
AD3084
dehydrogenase Atu4296 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD3084
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD3084
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AB008689; PIDN:AAL45090.1; PID:g17742757; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4296
A:Map position: linear chromosome

Query Match 4.2%; Score 122; DB 2; Length 527;
Best Local Similarity 19.8%; Pred. No. 0.23;
Matches 101; Conservative 55; Mismatches 175; Indels 180; Gaps 22;

QY 75 REFENLSABAVTRGVGMSTHWCSTPRHP---PMESLPGIGRPKLSNDPAEDDKWNE 131
Db 104 RDFREVEHE-----DGVSPAMPVSYAELEPPYAEARLFGV-RGRAGDDPTEPPRSAPY 156
QY 132 LYSEA--ERLIGTSKEPDESIRHTLVLSLQDAYKDRQIRFRPLA-----CHRL 181
Db 157 MHAPIHEPVGIRVAKGFER-----LGLRPFH-----MPSAIDYGPGLCRRC 199
QY 182 KNAPEYVWHSANLFSHSIYNDKOKKLF-----TLTNHRCRLALGGYKKIG 232
Db 200 GTCDAFVCFDAKG-----DAETRLRLPALRHPNVSLLTGARVRL--IADGGKHIV 250
QY 233 AAEVRLNLLATRNPPSOLDOSYIMAKYVVLASGAIGNPQILYNSGSLQVTPRNDSLIPN- 291
Db 251 AVEIER-----AGEITT-IEAPLFLVSAGAINSALLIRSA-----DEKKPNG 292
QY 292 -----LGRYTEQPM-----FCQIVLRQEFVDSVRDPP-----320
Db 293 LANSVGVGRYLMNHLGSLMGLLPFTINDTRFPKTMNSLNDFFDGTGDEAARGNVQMLG 352
QY 321 -----YGLFWKEAVAQHIKAKNPTDALPIFRDPEPQVTPPTTEHPWHTQIHRDAF 372
Db 353 NIQGPIMIRAAVPMWRPLANLLARHSVDFLVM-----SED-----387
QY 373 SYGAVGPEVDSRVIVDLRFWFGATDPPEANLLVFQNDVQGYSMPOPTFYRSTASN--- 429
Db 388 ----TPKYDSRV---KPGWKNAG-----LIYRPGDREAHQR 417
QY 430 -VRARKMMADMCEVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHMRSLLRKNGFPVVLGHSGFIEAPS-----HOCGTVMRGDDPKPEAALNALCOTY 469
```

Qy 489 DFANLYVAGNCTIRTFGENFTLSTMCCHAIK 519  
| | | | | : : : : :  
Db 470 DHPNLYVVDAGFFPSSAALNPALTAQAALR 500  
| | | | | : : : : :

## RESULT 14

T18562

hypothetical protein F15C11.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T18562; T20970

R:Barlow, K.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18976

A:Accession: T18562

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-502 &lt;W1L&gt;

A:Cross-references: EMBL:Z98262; PIDN:CA810932.1; GSPDB:GN00019; CESP:F15C11.2

A:Experimental source: clone VF15C11

R:Wilkinson, J.

submitted to the EMBL Data Library, April 1996

A:Reference number: Z19352

A:Accession: T20970

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-502 &lt;W12&gt;

A:Cross-references: EMBL:Z71260; PIDN:CAA95799.1; GSPDB:GN00019; CESP:F15C11.2

A:Experimental source: clone F15C11

C:Genetics:

A:Gene: CESP:F15C11.2

A:Map position: 1

A:Introns: 31/3; 67/1; 169/3; 212/3; 293/2; 294/1; 340/3; 368/3; 406/2; 428/2; 448/3; 48

Query Match 4.08; Score 116; DB 2; Length 502;

Best Local Similarity 21.98; Pred. No. 0.63;

Matches 110; Conservative 65; Mismatches 201; Indels 126; Gaps 26;

Qy 40 VSPVRNONTLPDGAWSAPPGSSAISNGKNPHOREPENLSABAVTHGVGSMSTHTWCS 99  
| | | | | : : : : :  
Db 74 VHLVIRNQARTPAPAA-ATTASAPSS--NP----- 103  
| | | | | : : : : :

Qy 100 TPRIHP-----PMBSLPGICGRP-KLSNDP-----AEDDKWENLYSEAE--RLIGTSTKE 146  
| | | | | : : : : :  
Db 104 TSSOPNTNPFPAAGMGSPADILNPNDAWSMDNPITQQLGNPEEFKRTIQSNPQ 163  
| | | | | : : : : :

Qy 147 F-----DESTRHTL-----VLSQDAYKDR---QRIFRPLPLACHRLKNAPEVEHSA 193  
| | | | | : : : : :  
Db 164 FOALIERNPEVGHILNDPNVMROTMEIRNPNMFEQEMMRNHQDAIRNIQGIPI-----GG 217  
| | | | | : : : : :

Qy 194 ENLFHSIYNDKQKLFLLTNHRCRLALRG-----GYEKKIGAAEVRNLLATR 243  
| | | | | : : : : :  
Db 218 EALERLYN-DVQPLNSATN-----SLSGNPFASLUGQSSEPRVDRAGQENNEALP 270  
| | | | | : : : : :

Qy 244 NPSSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTTPRNDLSILNGLRIITEQPMAF 303  
| | | | | : : : : :  
Db 271 NPWAS-----NANQATNQSNRSADFNLSLDSPGISLMQMSNPSNQASMF 319  
| | | | | : : : : :

Qy 304 COIVLRFQEFVSDRDPDGYLPMWKEAVAQHIKAP--TDAL--PIPRDPPEQVTTTFTE 359  
| | | | | : : : : :  
Db 320 -----SPEVINSIR-----QNMNNPGLIDSIVGQIPSAARDNPQISEGIRR 360  
| | | | | : : : : :

Qy 360 EHPWHTOIHDAFSYGAV-GPEVDSRVIVDLRFECATDPPEANNLL-VFQNDVQDGYSNPQ 417  
| | | | | : : : : :  
Db 361 SFQMLNMSDPSVMEAMRNPRVSEAFQIOEGFSTLRREAPQLNLNFQAGAMGGGAFGS 420  
| | | | | : : : : :

Qy 418 PTFRRYRSTASGNVARKMKDMCEVASNIGLYLTPSPQFMDP--GLALHLAGTTRIGF- 474  
| | | | | : : : : :  
Db 421 DA-----NASSAGANSACGLADLFN-SNMGGGRFSSTAAPVNPQTYASQLEQLQSMGFS 475  
| | | | | : : : : :

Qy 475 DKATTVDNNSLVWDFANLYYA 496

Db 476 DRARNVA---ALTATFGDLNAA 494  
| | | | | : | | | | |

## RESULT 15

B70917

probable zwf2 protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C&gt;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: B70917

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: B70917

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-514 &lt;COL&gt;

A:Cross-references: GB:Z95844; GB:AL123456; NID:G3250713; PIDN:CA809259.1; PID:G21310

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: zwf2

C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 4.08; Score 116; DB 2; Length 514;

Best Local Similarity 22.88; Pred. No. 0.66;

Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

Qy 29 FVNIKGAQQ-VSPVRNONTLPDGAWSAPPGSSAISNGKNPHOREPENLSAEAVTR 87  
| | | | | : | | | | |  
Db 77 FGQVYVNAVQEHCTPPFQQNWDRLAEG-FRFVGTGDTDDDAFAQLAETLEKLDRE---R 132  
| | | | | : | | | | :

Qy 88 GVGGMSTHTCSTPRIHPM-ESL--PGIGRPKLSNDPAEDDKWENLYSEAERLIGTST 144  
| | | | | : | | | | :  
Db 133 GTGGNHAFYLAIPPKSPFVCEQLHKSGLARPQ-----GDRWSRVVIE----- 175  
| | | | | : | | | | :

Qy 145 KEFDESIRHTLVLSQDAYKDRQIRPRLPLACHRL--KNAPEVVEHSAENLPHSIYN 202  
| | | | | : | | | | :  
Db 176 KPGHGLASARELNKAVNAVFPPEAVFR-----IDHYLGKETVQNILALRFANQLFDPIN 231  
| | | | | : | | | | :

Qy 203 ----DDKOKKLFLLTNHRCRLALTGGYEKKIGAAE--VRN-----LLATRNPSOL 249  
| | | | | : | | | | :  
Db 232 AHYVDHVQ-----ITMAEDIGLGRAGYDYGIGAARDVIONHLMQLLALTAMEEPVSFH 285  
| | | | | : | | | | :

Qy 250 DSYTMA-KVYVILASGAIGNP-----QILYNSGFSG----- 278  
| | | | | : | | | | :  
Db 286 PAALQAEKIKVLSATRLAEPLDQTTSRGOYAAAGWQGEKVVGLLDEEGFAEDSTTETFAA 345  
| | | | | : | | | | :

Qy 279 --LOVTPRNDSLIP-----NLGRYTEQPMAFQIVLRQEFV--DSVRDDPYGLPMWK 327  
| | | | | : | | | | :  
Db 346 ITLEVDTRRWAGVPFYLTGKRLGRVTEIALVF-----RRAPHLFPFDATMTDELG----- 396  
| | | | | : | | | | :

Qy 328 EAVAQHIKAKNPDTALPIPRDPPEQVTTTFTEHPWHQIHRDA---PSYGAV-----GPE 380  
| | | | | : | | | | :  
Db 397 -----TNAMVIRVQ-PDEGVTLRFSGKVPGTAMEVRDVMNMFESYSAFAEDSPE 444  
| | | | | : | | | | :

Qy 381 VDSRVIVDL 389  
| | | | | : | | | | :

Db 445 AYERLILDV 453  
| | | | | : | | | | :

Search completed: September 3, 2002, 15:58:48  
Job time: 533 sec

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:57:29 ; Search time 31.6 Seconds  
(without alignments)  
419.718 Million cell updates/sec

Title: US-09-856-327-2\_COPY\_76\_618  
Perfect score: 2901  
Sequence: 1 NAEETAVPVGYHKKNEI.....IINTLKGGTGKNTGEHRNL 543

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1142.5	39.4	623	1	US-08-734-925-2
2	1119	38.6	622	4	US-09-305-381-2
3	1024.5	35.3	586	4	US-09-023-731-1
4	192	6.6	51	4	US-09-023-731-12
5	159	5.5	615	4	US-09-297-937C-9
6	159	5.5	1276	4	US-09-297-937C-13
7	130.5	4.5	58	4	US-09-023-731-11
8	115	4.0	50	4	US-09-023-731-14
9	108	3.7	35	4	US-09-023-731-15
10	104.5	3.6	1876	2	US-08-609-049A-12
11	104.5	3.6	1876	4	US-09-170-996-12
12	102.5	3.5	61	4	US-09-023-731-6
13	96.5	3.3	1876	2	US-08-609-049A-28
14	96.5	3.3	1876	4	US-09-170-996-28
15	95.5	3.3	1257	1	US-08-049-783-2
16	95.5	3.3	1257	1	US-08-158-232-6
17	95.5	3.3	1257	1	US-08-304-626-6
18	95.5	3.3	1257	1	US-08-316-301A-6
19	95.5	3.3	1257	1	US-08-611-928-6
20	95.5	3.3	1257	3	US-09-173-891-6
21	95.5	3.3	1257	4	US-09-076-137-6
22	95.5	3.3	1257	5	PCT-US92-03624-6
23	95	3.3	3170	2	US-07-642-734C-5
24	95	3.3	3170	3	US-08-439-009A-5
25	92	3.2	2802	4	US-09-542-331-1
26	90.5	3.1	806	3	US-08-549-515-5
27	90.5	3.1	806	3	US-08-549-515-11

28 90 3.1 543 3 US-09-199-229-2 Sequence 2, Appli  
29 90 3.1 543 4 US-09-443-087-2 Sequence 2, Appli  
30 90 3.1 543 4 US-09-687-298-2 Sequence 2, Appli  
31 90 3.1 807 4 US-08-974-549A-5 Sequence 5, Appli  
32 90 3.1 1132 3 US-08-851-843A-225 Sequence 225, App  
33 90 3.1 1132 4 US-08-974-549A-2 Sequence 2, Appli  
34 90 3.1 1132 4 US-08-974-549A-344 Sequence 344, App  
35 90 3.1 1132 4 US-08-854-050-225 Sequence 225, App  
36 90 3.1 1132 4 US-09-430-323-225 Sequence 225, App  
37 90 3.1 1132 4 US-09-128-354-2 Sequence 2, Appli  
38 90 3.1 1154 4 US-08-974-549A-611 Sequence 611, App  
39 90 3.1 1189 4 US-08-974-549A-613 Sequence 613, App  
40 90 3.1 1200 4 US-08-974-549A-612 Sequence 612, App  
41 90 3.1 1285 4 US-08-974-549A-600 Sequence 600, App  
42 90 3.1 1407 4 US-08-974-549A-628 Sequence 628, App  
43 89 3.1 514 4 US-08-974-549A-605 Sequence 605, App  
44 88.5 3.1 1420 2 US-08-540-804-14 Sequence 14, Appli  
45 88.5 3.1 1420 2 US-08-218-265-14 Sequence 14, Appli

#### ALIGNMENTS

RESULT 1  
US-08-734-925-2  
; Sequence 2, Application US/08734925  
; Patent No. 5712139  
; GENERAL INFORMATION:  
; APPLICANT: NISHIMURA, IKUKO  
; APPLICANT: OKADA, KIMIHARU  
; APPLICANT: MINAMIHARA, TOMOYUKI  
; APPLICANT: KAWAI, GENSHIRO  
; APPLICANT: KOBAYASHI, YASUJI  
; APPLICANT: SUZUKI, MASARU  
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE  
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/734,925  
; FILING DATE: 22-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/568,428  
; FILING DATE: 06-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 7126-002-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 623 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-734-925-2

```
Query Match      39.4%; Score 1142.5; DB 1; Length 623;
Best Local Similarity 45.1%; Pred. No. 3.4e-112;
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

QY 13 GYHKNEIFQKIDIRFVNVIGALQOQVSPVVRNQNVPFLDPGANSAPGSSAISNGKNP 72
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 87 GAHKNTVEYQKIDKFVNVIQQLMSPVPVNTLVDTLSPASWQA--SSFFVRNGSNP 144
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 73 HOREFENISAEAVTRGVGGMSTHWTCTPRIHPPMESLPGIGRPKL-SNDPAEDDKENW 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 EODPLRNLSCQAVTRVVGGMSTHWTCTPRFEK-----LQRLVLKNDKADDAEDW 197
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 132 LYSEARLIGTSTKFEDESIRHTLVLSLQDAYKQRIFRPLPLACHRLKNAPEVVEWH 191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 LYTKAESYFKTCTDQKESIRHNLVNLKLAEEYKG-QRDFQOIPLAATR--RSPTFEVWS 254
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 192 SAENLF--HSIYNDKQKLFLLTNHRCRTRIALTGGYEKKIGAAEVRNLLATRNPSOL 249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 255 SANTVFDLQNRPNTPDAPNERFNLFFAVACERV-----VRN-----TSNSEI 295
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 250 DSY-----IMAKVYVLASGAINCPQILYNSGFSGLQVTPRND-----SLIPNLGRYITEQPMAP 303
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 296 ESHLHDLISGRFEKADVFVLTAGAVHNAQLLAAAGFQGL---GRDPDAKPLPSLLPYLGTHITEQTLVF 364
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 296 ITEQPMACQIVLROEFVDSVRDDPY--GLP-----WWKEAVAQH 333
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 356 ITEQSLVFCQVMTSELIDSVKSDMLIRGNPGDLGYSVYTPGAEKNKHPDMWNEKVKNH 415
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 334 IAKNTDALPIPRDPEQVTTPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFG 393
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 416 MNQHQEDPLPIPEPEQVTTLPQSPHWTQIHRDAFSYGAVQOQSIDSLRILVDWRFFG 475
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 394 ATDPANLLVQNDVQDGYSMQPTFRYR-PTASNVARKMMADMCEVASNLGYLPT 452
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 476 RETPEENKLFESDKITDYNMQPTDFRFPAGRTSKAEADMTDMCMVSAKIGGFPG 535
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 453 SPQPMDFGLALHLAGTTRIGFDKA--TVADNNSLVDFANLYVAGNGTIRTGFCENPT 510
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 536 SLQPMEPLGLVHLGGTHRMGFEDEQKCCVNTDSRVFGKFLGLGGCGNIPTAYGANPT 595
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 511 LTSMCHAKTSARSIIN 526
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 596 LTAMSALAKSCYEIKN 611
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
US-09-305-381-2
; Sequence 2, Application US/09305381
; Patent No. 6146865
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Flensted
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; TITLE OF INVENTION: Having Pyranose Oxidase Activity
; FILE REFERENCE: 5571.200-US
; CURRENT APPLICATION NUMBER: US/09/305.381
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/088,724
; EARLIER FILING DATE: 1998-06-10
; EARLIER APPLICATION NUMBER: PA 1998 00774
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Coriobacterium Versicolor
US-09-305-381-2

Query Match      38.6%; Score 1119; DB 4; Length 622;
Best Local Similarity 44.6%; Pred. No. 1.1e-109;
Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;

QY 13 GYHKNEIFQKIDIRFVNVIGALQOQVSPVVRNQNVPFLDPGANSAPGSSAISNGKNP 72
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 87 GSHKNTVEYQKIDKFVNVIQQLMSPVPVNTLVDTLSPASWQA--STFFVRNGANP 144
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 73 HOREFENISAEAVTRGVGGMSTHWTCTPRIHPPMESLPGIGRPKL-SNDPAEDDKENW 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 EODPLRNLSCQAVTRVVGGMSTHWTCTPRFEK-----LQRLVLKNDKADDAEDW 197
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 132 LYSEARLIGTSTKFEDESIRHTLVLSLQDAYKQRIFRPLPLACHRLKNAPEVVEWH 191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 LYTKAESYFKTCTDQKESIRHNLVNLKLAEEYKG-QRDFQOIPLAATR--OSPTFEVWS 254
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 192 SAENLF--HSIYNDKQKLFLLTNHRCRTRIALTGGYEKKIGAAEVRNLLATRNPSOL 249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 255 SAHTVFDLENRENPKDAPKQRFNLFFAVACTNVRNDRNANSEIVG-LDVRDLHGKGSIT--- 310
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 250 DSYIMAKVYVLASGAINCPQILYNSGFSGLQVTPRND-----SLIPNLGRYITEQPMAP 303
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 311 ---IRAKVITLTAGAVHNAQLLAAAGFQGL---GRDPDAKPLPSLLPYLGTHITEQTLVF 364
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 304 COIVLRQEFVDSVRDDP--YGLP-----WWKEAVAQHAKNPTDA 341
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 365 CQTVASTELINSVTADMTIVGKPGHPDYSVTVTPGNPNKHPDMWNEKVKKHMMDHQEDP 424
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 342 LPIDPRDPEQVTTPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDEANN 401
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 425 LPIDPRDPEQVTTLPQATHPWHTQIHRDAFSYGAVQOQSIDSLRILVDWRFFGRTPEKEEN 484
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 402 LVLVQNDVQDGYSMQPTFRYR-PTASNVARKMMADMCEVASNLGYLPTSPPOFMDPG 461
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 485 KLWFSKTKTDAYNLQPTDFR--FPGREAEDMTDMCMVSAKIGGFPGFPGMEPG 542
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 462 LALHLAGTTRIGFDKA--TVADNNSLVDFANLYVAGNGTIRTGFCENPTITSCHAIK 519
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 543 LVLHGLGTHRMGFEDEKADKCCVNTDSRVFGKFLGLGGCGNIPTAYANPTLTANSLAIK 602
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 520 SANSIINTLKGGTDCGNTGEHNRN 542
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 603 SCEYI---KKNEPSPNPVKHNN 622
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
```





REGISTRATION NUMBER: 33,285  
REFERENCE/DOCKET NUMBER: 3479-4000  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-023-731-11

Query Match 4.5%; Score 130.5; DB 4; Length 58;  
Best Local Similarity 44.4%; Pred. No. 1.5e-06;  
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;  
QY 258 YVLASGAIGNPQILYNSGSLQVTPRNDLSIPNLGRYTEQPMACQIVLRQEFVDSVR 317  
DB 2 FVIAGAVCTPQILWNS-----NIRPY-----ALGRYLSQSWTFQCIIVLKRGIYDAIA 50  
QY 318 DDP 320  
DB 51 TDP 53

RESULT 8  
US-09-023-731-14  
Sequence 14, Application US/09023731  
Patent No. 6291648  
GENERAL INFORMATION:  
APPLICANT: Kawamura, Yukio; Morita,  
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.  
TITLE OF INVENTION: ANTITUMOR PROTEIN AND  
TITLE OF INVENTION: GENE ENCODING SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,731  
FILING DATE: 13-FEB-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 29275/1997  
FILING DATE: 13-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KENNETH H. SONNENFELD  
REGISTRATION NUMBER: 33,285  
REFERENCE/DOCKET NUMBER: 3479-4000  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-023-731-14

Query Match 4.0%; Score 115; DB 4; Length 50;  
Best Local Similarity 53.5%; Pred. No. 5.2e-05;  
Matches 23; Conservative 5; Mismatches 15; Indels 0; Gaps 0;  
QY 409 VQDGYSPQPTFYRPSTASNVARKKMDMCEVASNLGGYLP 451  
DB 8 VTDYIGMPQPTFHVKRTNADGDRQRMNDMTNVANMLGGYLP 50  
RESULT 9  
US-09-023-731-15  
Sequence 15, Application US/09023731  
Patent No. 6291648  
GENERAL INFORMATION:  
APPLICANT: Kawamura, Yukio; Morita,  
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.  
TITLE OF INVENTION: ANTITUMOR PROTEIN AND  
TITLE OF INVENTION: GENE ENCODING SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,731  
FILING DATE: 13-FEB-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 29275/1997  
FILING DATE: 13-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KENNETH H. SONNENFELD  
REGISTRATION NUMBER: 33,285  
REFERENCE/DOCKET NUMBER: 3479-4000  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-023-731-15

Query Match 3.7%; Score 108; DB 4; Length 35;  
Best Local Similarity 66.7%; Pred. No. 0.00015;  
Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 453 SPQFMDPGLALHLAGTTRIGFDKATTVADNNS 485  
DB 2 SYQFMAPGLVHLITGTRIGTDDQTSVADPTS 34

RESULT 10  
US-08-609-049A-12  
Sequence 12, Application US/08609049A  
Patent No. 5948664  
GENERAL INFORMATION:  
APPLICANT: Williams, Lewis T.  
APPLICANT: Molz, Lisa  
APPLICANT: Chen, Yen-Wen  
TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides  
NUMBER OF SEQUENCES: 32

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, 8th Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/609,049A  
;; FILING DATE: 29-FEB-1996  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dow, Karen B.  
;; REGISTRATION NUMBER: 29,684  
;; REFERENCE/DOCKET NUMBER: 2307K-063700US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-326-2400  
;; TELEFAX: 415-326-2422  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1876 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-609-049A-12

Query Match 3.6%; Score 104.5; DB 2; Length 1876;  
Best Local Similarity 22.0%; Pred. No. 0.37; Indels 79; Gaps 21;  
Matches 85; Conservative 53; Mismatches 170;  
QY 8 VPYVGYHKKNEIEFKDIDREVNVIKALQOVSVVPRNQNVPTLDPGA---WSAP--P 61  
DB 184 VPYQAAQQOORPLN--SEELQRLYSM---PAQMAVVVPQPNAYMYYPGAVVTPYTAIPV 239  
QY 62 GSSAITSNKNPHQREFENLSAEAVTRGVGGMSTHTCTSPRIHP-PMESLPGIG--RPKL 118  
DB 240 GSAFMPFPQYPAQ-----GYGFGGAYTHMDLRRPQSQAPQQTATPTTSHHSQP 288  
QY 119 SNDPAEDDKENNELYSEAE-----LIGTSTKEFDESIRHTLVRLSLQDA---YKDRQRI 170  
DB 289 SNHSTSSPAEANGVAFAPARRQVPSTVGVSSTHTGNNGHSSVPRRGNDLIDLNHEDYSRV 348  
QY 171 -----FRPLPLACHRLKNAPEYVEWHSANLPHSYIYNDKQKKLFTLLTNHRCRTRALGTG 225  
DB 349 SVLEAFDPLNDNTGNDTASDSTSYAEVDPDFLYSGD-----AATQYSDPMYEA 401  
QY 226 GYEKKIGAEVR-NLLATR-----NPSQOLDYSYMA-----KYVVLASGAIGNP----- 268  
DB 402 RWDKTAVATVSPNVGLIGRWQDFLSQPSSTSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461  
QY 269 --QILYNSGFGLOVT-PRNDSLIPNLGRYTEQPMACQIVLRO-----EFVDSVR 317  
DB 462 RNQCYESNQAAAMPVSRPPQSSVLTD--SYTSSIP---ANVVLDRRKCTRLYELISDQR 516  
RESULT 11  
US-09-170-996-12  
; Sequence 12, Application US/09170996  
; Patent No. 6291220  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Molz, Lisa

;; APPLICANT: Chen, Yen-Wen  
;; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides  
;; NUMBER OF SEQUENCES: 32  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, 8th Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/170,996  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/609,049  
;; FILING DATE: 29-FEB-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dow, Karen B.  
;; REGISTRATION NUMBER: 29,684  
;; REFERENCE/DOCKET NUMBER: 2307K-063700US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-326-2400  
;; TELEFAX: 415-326-2422  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1876 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-170-996-12

Query Match 3.6%; Score 104.5; DB 4; Length 1876;  
Best Local Similarity 22.0%; Pred. No. 0.37; Indels 79; Gaps 21;  
Matches 85; Conservative 53; Mismatches 170;  
QY 8 VPYVGYHKKNEIEFKDIDREVNVIKALQOVSVVPRNQNVPTLDPGA---WSAP--P 61  
DB 184 VPYQAAQQOORPLN--SEELQRLYSM---PAQMAVVVPQPNAYMYYPGAVVTPYTAIPV 239  
QY 62 GSSAITSNKNPHQREFENLSAEAVTRGVGGMSTHTCTSPRIHP-PMESLPGIG--RPKL 118  
DB 240 GSAFMPFPQYPAQ-----GYGFGGAYTHMDLRRPQSQAPQQTATPTTSHHSQP 288  
QY 119 SNDPAEDDKENNELYSEAE-----LIGTSTKEFDESIRHTLVRLSLQDA---YKDRQRI 170  
DB 289 SNHSTSSPAEANGVAFAPARRQVPSTVGVSSTHTGNNGHSSVPRRGNDLIDLNHEDYSRV 348  
QY 171 -----FRPLPLACHRLKNAPEYVEWHSANLPHSYIYNDKQKKLFTLLTNHRCRTRALGTG 225  
DB 349 SVLEAFDPLNDNTGNDTASDSTSYAEVDPDFLYSGD-----AATQYSDPMYEA 401  
QY 226 GYEKKIGAEVR-NLLATR-----NPSQOLDYSYMA-----KYVVLASGAIGNP----- 268  
DB 402 RWDKTAVATVSPNVGLIGRWQDFLSQPSSTSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461  
QY 269 --QILYNSGFGLOVT-PRNDSLIPNLGRYTEQPMACQIVLRO-----EFVDSVR 317  
DB 462 RNQCYESNQAAAMPVSRPPQSSVLTD--SYTSSIP---ANVVLDRRKCTRLYELISDQR 516  
RESULT 12

```

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSES: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%: Score 102.5; DB 4;
Best Local Similarity 43.5%; Pred. No. 0.0016;
Matches 27; Conservative 6; Mismatches 24;

QY 29 FVNVIKALQVYVPRNQNVPTLDGAWSPGSSA----ISNG
      ||| ||| |: : ||| || ||| :
Db 1 FVNIINGALQPTISPSDTYQPTLAVAAW-APPIDPAGQLVING
      ||
QY 85 VT 86
      ||
Db 60 VT 61

RESULT 13
US-08-609-049A-28
; Sequence 28, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664e1 PI 3-Kinase Pol
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

```

```

; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-609-049A-28

      Query Match          3.3%; Score 96.5; DB 2; Length 1876;
      Best Local Similarity 21.7%; Pred. No. 2.6;
      Matches 84; Conservative 53; Mismatches 171; Indels 79; Gaps

Qy  8  VPVVPGVGHKKEIEFQKIDRFYNNVIKALQQVSVPRNQNVPTLDPGA----WSAP--p 61
Db  184 VPIQPAQQQRPLN-SEELQRLYSM---PAQMAVVPVQPQPNAYWYPGAVVTPYTAIYP 239
Qy  62  GSSAISNGKPHOREFNLSAEAVTRGVGGMSTHTWCSTPRIHP-PMESLPGIG--RPKL 118
Db  240  GSAAFMPPQYPAQ-----GYGFGGAYTHMDLRRPQSQAPAQQTATTSHHHSQP 288
Qy  119 SNDPADDEKKNELYSAEK---LIGTSKFEDESIRHTLVLSLQDA----YKDRQRI 170
Db  289 SNHSTSSPAEANGVAPPARQVSTVGVSSSSITGNHGSSVPRRQNDLIDLNHEDYSRV 348
Qy  171 -----FRPLPLACHRLKNAPEYVEVHSAENLFHSIYNDOKKLEFLLTNHRCRLALTG 225
Db  349 SVLEAFDPLNDNTGNDTSDSYSAEYDPDFLYSGD-----AATQYSDPMYEAVN 401
Qy  226 GYEKKIGAAEVR-NLIATR-----NPSSQLDSITMA-----KVYVLASGATGNP----- 268
Db  402 RWDKTVATVSPNVNLIGWRQDFLSQPSSTSSQYGVAPPEESLKLAEENGSETISPPPLPP 461
Qy  269 --QILYNSGSGLOVT-PRNDSLIPNLGRVITQPMAFCQIVLRQ-----EFVDSVR 317
Db  462 RNQOCYESNQAAAMPVSRPPOSSVLTD--SYTSSIP---ANVVLDRRKCTRILYELISQOR 516
Qy  318 -DDPYGLPWWKEAQAQHIKAKNPDTDALP 343
Db  517 TDDELLEFFV-HMVKVEARYPHDDAP 542

RESULT 14
US-09-170-996-28
; Sequence 28, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220e1 PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco

```





Db	625	-----QDSPPEVHESTIIIDKSPPTIWSSNKHSHYSHIHLEG-SYTSQG-SYPHNLLINL	676
Qy	390	RWFGATDPEANNLLVFON-DVQDGYSMPOPTFRYRPTASNVRRKMMADMCEVASNLG-	447
Db	677	--FHPTDPRNRNHTIHVWNGDMNVYD-----KDSVADGLNFNKITATIPSDAWYSGT	726
Qy	448	-----GYLPTSPQFMDPGLALHLAG-TTRIGFDKAT-----TVADNNSLWDFANL	493
Db	727	ITSMHLEFNDNFKTITPKF---ELSNLENIITTQVNALFASSAQDTILASNVSDYW-IEQV	782
Qy	494	YVAGNGTIRTGFG-ENPTLTSMCHAISARSIIINTLKGG	531
Db	783	VMKYDALSDVEFGKEKKALRKLNVQAKRLSKIRNLLIGG	821

Search completed: September 3, 2002, 15:57:34  
Job time: 3144 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:56:34 ; Search time 82.18 Seconds  
(without alignments)  
733.914 Million cell updates/sec

Title: US-09-856-327-2\_COPY\_76\_618

Perfect score: 2901

Sequence: 1 NAEETAVPVPGYHKKEI.....INTLKGTDGKNTGEHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2901	100.0	618	22	Lyophyllum shimeji
2	1142.5	39.4	623	20	Coriolus versicolor
3	1136	39.2	633	21	Pleurotus cornucop
4	1128.5	38.9	623	17	Pyranose oxidase
5	1119	38.6	622	22	Trametes hirsuta p
6	1024.5	35.3	566	20	An antitumour prot
7	976.5	33.7	564	21	T. matsutake pyran
8	976.5	33.7	564	21	Trichoderma deri
9	192	6.6	51	20	Peptide derived fr
10	177	6.1	67	21	T. matsutake pyran
11	177	6.1	67	21	N-terminal fragmen

12	168	5.8	34	22	AA97034	Lyophyllum shimeji
13	162	5.6	30	22	AA97031	Lyophyllum shimeji
14	159	5.5	615	20	AA905719	Gluconate dehydrog
15	157	5.4	544	20	AA914049	G. oxydans D-sorbi
16	142	4.9	29	22	AA97033	Lyophyllum shimeji
17	130.5	4.5	58	20	AA987540	Peptide derived fr
18	128.5	4.4	657	21	AA915525	Arabidopsis thalia
19	128.5	4.4	675	21	AA915524	Arabidopsis thalia
20	128.5	4.4	748	21	AA915523	Arabidopsis thalia
21	121	4.2	24	22	AA97032	Lyophyllum shimeji
22	115	4.0	50	20	AA987543	Peptide derived fr
23	114.5	3.9	704	20	AA931754	Candida tropicalis
24	110	3.8	551	21	AA913598	Streptomyces globi
25	108	3.7	35	20	AA987544	Peptide derived fr
26	106.5	3.7	778	20	AA935090	Chlamydia pneumoni
27	106	3.7	546	20	AA928581	Choline oxidase (C
28	105	3.6	609	22	AA956302	Propionibacterium
29	105	3.6	833	22	AB968116	Drosophila melanog
30	104.5	3.6	1356	22	AB967291	Drosophila melanog
31	104.5	3.6	1876	18	AA938757	Phosphatidyl inosi
32	104.5	3.6	1876	22	AB961970	Drosophila melanog
33	104.5	3.6	1876	22	AB966876	Drosophila melanog
34	103	3.6	437	22	AB929360	Novel human diagno
35	103	3.6	1688	22	AB959327	Drosophila melanog
36	102.5	3.5	61	20	AA987535	Peptide derived fr
37	102.5	3.5	352	22	AB971452	Drosophila melanog
38	102.5	3.5	1686	19	AA970991	Human class II P13
39	102	3.5	1130	20	AA913446	Human aortic carbo
40	102	3.5	1244	22	AB957666	Human protein kina
41	100.5	3.5	880	22	AB971719	Drosophila melanog
42	100	3.4	522	22	AB958468	Drosophila melanog
43	100	3.4	522	22	AB958474	Drosophila melanog
44	100	3.4	645	21	AA941556	Human ORFX ORF1320
45	98.5	3.4	777	22	AB919089	Novel human diagno

#### ALIGNMENTS

RESULT 1

AA97035

ID AA97035 standard; Protein; 618 AA.

XX AC AA97035;

DT 20-JUL-2001 (first entry)

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

XX DE Lyophyllum shimeji

PT cost on large scale -  
XX Claim 3; Page 43-45; 52pp; Japanese.  
XX  
XX The present sequence is an antibacterial protein from the fungus  
CC *Lyophyllum shimeji*. The protein was obtained from a fraction prepared  
CC by extracting *Lyophyllum shimeji* with water and inhibiting the extract  
CC to ammonium sulphate precipitation. The protein inhibits the growth  
CC of the plant pathogenic bacteria *Pseudomonas solanaceae* and *Rhizoctonia*  
CC *solanii* at a relatively low concentration. P. orizae and R. solani are  
CC causative of the two major diseases of rice. The protein contains  
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.  
XX The antibacterial protein can be produced at low cost on a large scale.  
XX  
SQ Sequence 618 AA;  
  
Query Match 100.0%; Score 2901; DB 22; Length 618;  
Best Local Similarity 100.0%; Pred. No. 5.1e-266;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NAEGETAVPYPCYHKKKIEFQKIDRFVNVKIGALQOVSVPRNVNPTLDPGAWSAP 60  
Db |||||  
QY 76 NAEGETAVPYPCYHKKKIEFQKIDRFVNVKIGALQOVSVPRNVNPTLDPGAWSAP 135  
Db |||||  
QY 61 PGSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKLSN 120  
Db |||||  
QY 136 PYSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKLSN 195  
Db |||||  
QY 121 DPAEDDKENNELYSEARLIGTSTKEFDESIRHTLVRLSLQDAYKDRQIRFRLPLACHR 180  
Db |||||  
QY 196 DPAEDDKENNELYSEARLIGTSTKEFDESIRHTLVRLSLQDAYKDRQIRFRLPLACHR 255  
Db |||||  
QY 181 LKNAPYVWHSANLPHSIYNDKQKLFLLTNHRCRTRALTGSEYKKGIGAAEVRNLL 240  
Db |||||  
QY 256 LKNAPYVWHSANLPHSIYNDKQKLFLLTNHRCRTRALTGSEYKKGIGAAEVRNLL 315  
Db |||||  
QY 241 ATRNPSQLDSYIMAKVYVLSAGSAGNPOILYNSGFSGLQVTPRNDLSLIPNLGRYTEQP 300  
Db |||||  
QY 301 MAFCQIVLRQEVDSVRDPPYGLPWKKEAVAOHIAKNPTDALPIPRDPPEQVTPPTTEE 360  
Db |||||  
QY 376 MAFCQIVLRQEVDSVRDPPYGLPWKKEAVAOHIAKNPTDALPIPRDPPEQVTPPTTEE 435  
Db |||||  
QY 361 HPWHTQIHRDAFSYGAVGPEVDSRVLDLRFWFGATDPPEANLLVFQNDVDGYSMPQPTF 420  
Db |||||  
QY 436 HPWHTQIHRDAFSYGAVGPEVDSRVLDLRFWFGATDPPEANLLVFQNDVDGYSMPQPTF 495  
Db |||||  
QY 421 RYRPTASNVARKMADMCCEVASNLGGYLPTSPQFMDPGLALHAGTTRIGFDKATTV 480  
Db |||||  
QY 496 RYRPTASNVARKMADMCCEVASNLGGYLPTSPQFMDPGLALHAGTTRIGFDKATTV 555  
Db |||||  
QY 481 ADNLSLWDFANLYVAGNTITRTGGENPTLTSCHAKSARSIINTLKGKDGKNTGEH 540  
Db |||||  
QY 556 ADNLSLWDFANLYVAGNTITRTGGENPTLTSCHAKSARSIINTLKGKDGKNTGEH 615  
Db |||||  
QY 541 RNL 543  
Db |||||  
QY 616 RNL 618  
Db |||||  
  
RESULT 2  
ID AAW94308  
XX AAW94308 standard; protein; 623 AA.  
AC AAW94308;  
XX  
XX 08-APR-1999 (first entry)  
XX  
XX Coriolus versicolor pyranose oxidase.  
DE  
XX Coriolus versicolor; modified; pyranose oxidase; heat stability;  
KW

KW pH stability; glucose; glucosone.  
XX Coriolus versicolor.  
XX JP11009271-A.  
XX 19-JAN-1999.  
XX 24-JUN-1997; 97JP-0181865.  
XX 24-JUN-1997; 97JP-0181865.  
XX (KIKK ) KIKKOMAN CORP.  
XX WPI; 1999-145891/13.  
XX  
XX New modified pyranose oxidase gene and protein - useful for  
PT efficient recombinant production of pyranose oxidase with excellent  
PT Km, heat and pH stability  
XX  
PS Claim 1; Page 7-9; 10pp; Japanese.  
XX  
CC The present sequence represents pyranose oxidase derived from *Coriolus*  
CC versicolor, which can be modified by adding, deleting or replacing at  
CC least one amino acid (preferably where the amino acid at position 542  
CC is replaced) and still retains pyranose oxidase activity. The modified  
CC pyranose oxidase with the amino acid at position 542 replaced, has the  
CC following properties: (a) Action: It oxidises glucose to glucosone;  
CC (b) Substrate specificity: It reacts specifically with glucose and also  
CC reacts with galactose, L-sorbose, D-xylose and L,5-anhydro-D-glucitol;  
CC (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)  
CC Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;  
CC and (f) Temperature stability: Stable up to 55 degrees Celsius. The  
CC present invention also describes a method for the preparation of a  
CC modified pyranose oxidase. The method efficiently prepares a modified  
CC pyranose oxidase with excellent Km value, heat stability and  
XX pH stability.  
XX  
SQ Sequence 623 AA;  
  
Query Match 39.4%; Score 1142.5; DB 20; Length 623;  
Best Local Similarity 45.1%; Pred. No. 4.1e-99;  
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;  
  
QY 13 GYHKKNELEFQKIDRFVNVKIGALQOVSVPRNVNPTLDPGANSAPGSSAISNGRNP 72  
Db |||||  
QY 87 GAHKKNTVEYQKIDRFVNVKIGALQOVSVPRNVNPTLDPGANSAPGSSAISNGRNP 144  
Db |||||  
QY 73 HQREFENLSAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKLSNDPAE--DKKEWNE 131  
Db |||||  
QY 145 EQDPIRLNSGQAVTRVGGMSTHTWCATPRFREQ-----RPLVKKDDQDADAEDWR 197  
QY 132 LYSEARLIGTSTKEFDESIRHTLVRLSLQDAYKDRQIRFRLPLACHRLKNAPYVEWH 191  
Db |||||  
QY 198 LYTKAESYKIGTDDGFKESIRHNLVNLKLAEEYKG--QDFGQIPLAATR--RSPTFVWS 254  
QY 192 SAENLF--HSIYNDKQKLFLLTNHRCRTRALTGSEYKKGIGAAEVRNLLATRPNSQL 249  
Db |||||  
QY 255 SANTVIDLQNRPTDAPNERINLPVACERV-----VRN-----TSNSEI 295  
QY 250 DSY-----IMAKVYVLSAGSAGNPOILYNSGFSGL--QVTPRNP--DSLIPNLGRY 295  
Db |||||  
QY 296 ESLHIDILISGDRFEIKADVLTAGAVHNAQLVNSGFGQLGRPDPAHPQLPSLSGY 355  
QY 296 ITEQPMACQIVLRQEVDSVRDPPY--GLP-----WWKEAVAOH 333  
Db |||||  
QY 334 TAKNPTDALPIPRDPPEQVTPPTTEEHPWHTQIHRDAFSYGAVGPEVDSRVLDLRFWG 393  
Db |||||  
QY 416 MMHQEDQPLPFPEDPEPQVTLTFQSPHWHTQIHRDAFSYGAVGQSDLSRLIVDWRFFG 475  
Db |||||

```
QY 394 ATDPANNLLVFQNDVODGYSMFQPTFRYR--PSTASNVARKMADMCCEVASNIGGYLPT 452
Db 476 rtepkeenklwfsdkitdympqptfdfrpgrtskaedmntdmcvmsakigglpg 535
QY 453 SPQFMDPGLALHLAGTTRIGFKA--TTVDANNSLVWDFANLYVAGNGTIRTGFGENPT 510
Db 536 slpqfmeqglvhlhgghrmgfeqekdccevnldsrvgfknlfiggcgnlptayganpt 595
QY 511 LTSMCHAIKSARSIIIN 526
Db 596 ltamslaikaceyikn 611

RESULT 3
AAV52700
ID AAV52700 standard; Protein; 633 AA.
XX
AC AAV52700;
XX
DT 07-MAR-2000 (first entry)
XX
DE Pleurotus cornucopiae antitumour protein.
XX
KW Antitumour; cancer; tumour; treatment; expression; tumour suppressor;
KW p53; pBR; ss.
XX
OS Pleurotus cornucopiae.
XX
PN JP11315096-A.
XX
PD 16-NOV-1999.
XX
PF 07-AUG-1998; 98JP-0236349.
XX
PR 08-AUG-1997; 97JP-0215311.
PR 02-MAR-1998; 98JP-0066176.
XX
PA (NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.
XX
WPI; 2000-058170/05.
DR N-PSDB; AA246411.
XX
An antitumour protein derived from Pleurotus cornucopiae and its gene -
useful for treatment of cancer including those caused by abnormal
expression of cancer inhibitory gene (e.g. p53 and pBR)
XX
Claim 3; Page 13-15; 23pp; Japanese.
XX
The invention relates to a novel antitumour protein extracted from
fruiting bodies of the fungus Pleurotus cornucopiae. The protein and
nucleotides encoding it are useful for the treatment of cancer,
including those caused by abnormal expression of tumour suppressor
genes such as p53 and pBR. This sequence represents the antitumour
protein.
XX
SQ Sequence 633 AA;

Query Match 39.2%; Score 1136; DB 21; Length 633;
Best Local Similarity 45.3%; Pred.No. 1.8e-98;
Matches 248; Conservative 70; Mismatches 175; Indels 54; Gaps 15;

QY 13 GYHKKEIEFKQKIDREVNVIKALQOVSVFVRNQNVPTLDPGAWSPGSSAISNGKNP 72
Db 96 gshkktveyqknidkfvhqqgmpvsvpnkvyadvdtlspasqa--sthfvrganp 153
QY 73 HQREFENLSAEAVTRGVGGMSTHWTCSTPRIHPHMESLPGIGRPKL-SNDPAEDDKWNE 131
Db 154 eqdpfnlsggavtrvggmthwtcatprfhks-----erpklvkdssadalewer 206
QY 132 LYSEAEELICTSKEPDESIRHVLRLSDAYKDRQRIERPLPLACHRLKNAPEYVEWH 191
Db 207 lydlaesfvktghnqfqsirhlvleklqdesyag-grgfeqiplaaqrtn--prfvws 263
```

```
QY 192 SARNLF--HSIYNDKQKKLFTLLTNHRCRLALTGGYERKKIGAAAEVNRLLATRNPSQL 249
Db 264 sahtvfdlenrpnaddekgrfnlpavvcervt-rdsldrkiedievhdliag----- 315
QY 250 DSY-THAKVTVVLASGAIGNPQILYNSGFSGLQVTPRND-----SLIPNLGRYTEQPMAPC 304
Db 316 drykvkadvfilcsgavhnpqilvnsqf-grmgqpdsslppttllpylsviteqtlftc 374
QY 305 QIVLROEFVDSVRDDP--YGLP-----WWKEAVAQHIKAKNPTDAL 342
Db 375 qtvfstelvnlvksdmliivgtpgqdyvstftdpdenkbpnwnnekvqhmhmqhqdpl 434
QY 343 PIPRDPPEQVTTPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFEGATDPEANNL 402
Db 435 piplddpepqvttlfqthpwtqihrdafsygavaesidrlvvdwrfgrtepvveen 494
QY 403 LVFQNDVODGYSMFQPTFRYR--PSTASNVARKMADMCCEVASNIGGYLPTSPQFMDPG 461
Db 495 lwfskqitdaynlpqtfstrfpqgrtaqeaelmmadmctmstkvggflpgsyppqfmapg 554
QY 462 LALHLAGTTRIGFDKATTV--DNNSLVWDFANLYVAGNGTIRTGFCENPTLTSMCHAIK 519
Db 555 lvhlqgthmgfdeeadkacvdtnskvgmenlfiggcgnigtayasnptltcavalair 614
QY 520 SARSIIIN 526
Db 615 sckyirn 621

RESULT 4
AAR9628
ID AAR9628 standard; Protein; 623 AA.
XX
AC AAR9628;
XX
DT 27-NOV-1996 (first entry)
XX
DE Pyranose oxidase.
XX
KW Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
KW marker; diagnosis; 1,5-anhydro-D-sorbitol.
XX
OS Coriolus versicolor.
XX
PN DE19545780-A1.
XX
PD 13-JUN-1996.
XX
PF 07-DEC-1995; 95DE-1045780.
XX
PR 24-MAY-1995; 95JP-0124835.
PR 07-DEC-1994; 94JP-0304086.
XX
PA (KIKK ) KIKKOMAN CORP.
XX
PI Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
PI Suzuki M;
XX
WPI; 1996-278990/29.
DR N-PSDB; AAT34420.
XX
DNA encoding protein with pyranose oxidase activity at neutral pH -
useful for the determination of glucose in body fluids or foods, or
1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis
XX
Claim 1; Page 14-16; 22pp; German.
XX
The present sequence encodes a protein isolated from Coriolus versicolor,
which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
glucose to glucosone and has an optimum pH of 7-7.5. It has a mol. wt. of
290000 (determined by gel filtration) and is stable at around 50deg.C.
PO can be used for measurement of glucose in, e.g. foods or body fluids,
```



```
OY 520 SARSIIINTLKGGTGDKNYGEHRN 542
Db 603 sceyi-----kknfepnnpvkhn 622

RESULT 6
AAW87531
ID AAW87531 standard; Protein; 566 AA.
AC
XX AAW87531;
XX
XX 26-FEB-1999 (first entry)
DE An antitumour protein.
XX
XX Antitumour protein; Tricholoma matsutake.
XX
XX Tricholoma matsutake.
XX
XX JP10313876-A.
XX
XX 02-DEC-1998.
XX
XX 13-FEB-1998; 98JP-0031452.
XX
XX 13-FEB-1997; 97JP-0029275.
XX
XX (MOMO-) MOMOYA KK.
PA (NORO-) NORINSUISANSO SHOKUHN SOGO.
XX
XX WPI; 1999-074153/07.
DR N-PSDB; AAW83626.
XX
XX An anti-tumour protein - prepared by culture of host cell
PT transformed by vector containing base coding sequence
PT
XX Claim 1; Page 7-8; 15pp; Japanese.
XX
XX The present sequence encodes an antitumour protein, and is isolated
CC from Tricholoma matsutake.
XX
XX Sequence 566 AA;

Query Match 35.3%; Score 1024.5; DB 20; Length 566;
Best Local Similarity 43.0%; Pred. No. 5.6e-88;
Matches 232; Conservative 77; Mismatches 180; Indels 51; Gaps 12;

OY 2 ABECTAVPVYGYHKKKEIEFQKIDRFVNVVIGALQOVVFNQVNPVILDPGAWGAPP 61
Db 62 aeiqsgdnpvigaahkksikfkdkfvnlngalqpsisdytqptlavaaw-app 120
OY 62 GSSA-----ISNGKNPHOREFENLSAEAVTRGVGGMSTHWCTSTPHIHPMESLPGIGRPK 117
Db 121 idpaegqlvlnghnpdeaglnlpgsavtrvvggmachwlcacptphd-----ee 170
OY 118 LSNDPAEDKEMNELYGEARLIGTSTKEFDESIRHTLVLRSLQADYKDRQIRPRPLA 177
Db 171 rvnnpv-dkqefdalleraktlnlvhdsqdydsirgqvketlqgtl-dasrgvttlpg 228
OY 178 CHRUKNAPEVWEHSAENLFSIYNDQKQKFLTLNHRCTRUALGGYEKKTGAAEVR 237
Db 229 verrtdnpiyvtwagad-----tvlgdvpkspfrfalvtetrvtklivsetnptqvvaallr 284
OY 238 NLLATRPNSOLDYSYINAKVYVLASGAIGNPQIILYNSGFSGLQVTPRNDLSLIPNLGRYIT 297
Db 285 nl-----ntsn--delvaksfviacgavctpqilwns-----nlrpy-----algryls 327
OY 298 EQPMAFQCIIVLRQEFVDSVRDDPYGLPWPWKEAVAQHIKAKNPTDALPIPFRRDPEQVTTTF 357
Db 328 eqsmtfcqilvkgivdaiaidpr-----faakveahkkkbpddvlpipfhepeqvmipy 383
OY 358 TEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFAGATDPANNLLVF-----QND 408
```

```
Db 384 tsdfpwhvqhrdafsygvgpkadprvvvdrlrfgksdiveenrvtfgpnpkIreweag 443
OY 409 VQDGYSMQPQTFERYRPTASNVARKKMDMCEVASNLGGVLTSPFPQFMDPGALHLAG 468
Db 444 vtdtygmppqptfhwkrtnadgdrqmmndmtnvnmllggylpqsyqqfnapglvlhitg 503
OY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGENTPTLTSMCHAKSARSIIINTL 528
Db 504 ttrigtddqtsvadptskvhnfnlwgvgngcipcdatacnprrtsvayalkgaeavvnyl 563

RESULT 7
AAB10457
ID AAB10457 standard; Protein; 564 AA.
XX
XX AAB10457;
XX
XX 11-DEC-2000 (first entry)
DE
XX T. matsutake pyranose oxidase protein.
XX
XX Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
XX diabetes mellitus.
XX
XX Tricholoma matsutake.
XX
XX JP2000175698-A.
XX
XX 27-JUN-2000.
XX
XX 16-DEC-1998; 98JP-0357423.
XX
XX 16-DEC-1998; 98JP-0357423.
XX (NISB) JAPAN TOBACCO INC.
XX
XX WPI; 2000-478485/42.
DR N-PSDB; AAA71487.
XX
XX A reagent containing pyranose oxidase for the determination of pyranose
PT
PT Claim 5; Page 17-18; 23pp; Japanese.
XX
XX This invention describes a novel reagent containing pyranose oxidase for
CC the determination of pyranose which can be prepared from a fraction
CC precipitated from an aqueous extract of matsutake mushroom by ammonium
CC sulfate precipitation and has an antibacterial activity against at least
CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
CC 210 kD by gel filtration and shows the presence of components of ca.
CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
CC activity by being heated at 60 degrees C for 10 minutes in a neutral
CC aqueous solution and in which the above antibacterial activity is
CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
CC aqueous solution. The invention also describes a method for the
CC determination of pyranose in a sample in which the above reagent for the
CC determination of pyranose is reacted with pyranose in the sample and the
CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
CC method for diagnosing a disease accompanied by abnormality in sugar
CC metabolism in which the above reagent for the determination of pyranose
CC is reacted with pyranose in the sample and the hydrogen peroxide formed
CC is reacted with an enzyme to develop a color, and a kit for pyranose
CC analysis or the diagnosis of a disease accompanied by abnormality in
CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
CC for diabetes mellitus. This sequence represents the Tricholoma matsutake
CC pyranose oxidase protein which is described in the method of the
XX invention.
XX Sequence 564 AA;

Query Match 33.7%; Score 976.5; DB 21; Length 564;
```

Best Local Similarity 41.3%; Pred. No. 2e-83;  
Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;

QY 2 AEEGTAVPVVPGYHKKNETEFKIDRFVNVKLGALQOVSVPRNQNVPTLDPGAWSAAPP 61  
DB 62 aelsgqdnvlgahhrnsikfqtdktdkfvniingalqpsispsdyqptlavaaw-app 120  
QY 62 GSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHWTCTSPRIHPMESLPGIGRPK 117  
DB 121 idpaegqlvimhnpqeaqlnlpgsavtrvgmawhwtcaactphd-----ee 170  
QY 118 LSNDAEDDKENNELYSAEERLIGTSTKEFDESIRHTLVLSQDAYKDRQRIPLPLA 177  
DB 171 rvnnpv-dkqefdalleraktllnvhsdgydsirgqvketlqtl-dasrgvtlplg 228  
QY 178 CHRLNAPYEVVHSAENLFHSIYNDKQKLFLLTNHRCRLALTGGYEKKIGAAEVR 237  
DB 229 verrtdnplyvtwtgad---tvlgdvkpsrfrvltvtrvktfivsetnptqvvaalir 284  
QY 238 NLLATRNPSQLDSYIMAKVYVYVLAIGAIGNPQILYNSGFSGLQVTPRNDSLIPNLGRYIT 297  
DB 285 nl-----ntsn--delvvaqsviaagavctpqilwns-----nirph-----algryls 327  
QY 298 EQPMAPFCQIVLQREFVDSVRDDPYGLPWKEAVAQHIKKNPTDALPIPRDPPEQVTPFP 357  
DB 328 eqsmtfcqivlkrsvdsiatdpr----faakveahkkhpdvlpiphepeqvmipy 383  
QY 358 TTEHPWHTQIHRDAFSYGAAGPEVDSRVIVDLRWFGADPEANLLVF-----QND 408  
DB 384 tsdfpwhvqvr--yafgdvqpkadprvvvdrlfrfgksdiveenrvtfgnpkldweag 441  
QY 409 VODGYSMPQPTFRYRSTASNVARKMMADMCEVASNLGGYLTSPQPMDFGLALHLAG 468  
DB 442 vtdtygmpeptfhvkrtnadgrdqrmdmtnvniigylpgsyppqfmapglaqhitg 501  
QY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGIRTFGFGENPFLTSMCHAIKARSIIINTL 528  
DB 502 ttrigtddqtsvadptskvhnfdnlwvgngcipcdatacnprtstsvayalkgaeavvyl 561

## RESULT 8

AAW81952  
ID AAW81952 standard; Protein; 564 AA.  
XX  
AC AAW81952;  
XX  
DT 07-JUL-2000 (first entry)  
XX  
DE Trichoderma derived antifungal protein sequence.  
XX  
KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;  
KW growth inhibitor; plant pathogenic fungi; antibacterial agent;  
KW N-terminal fragment.  
XX  
OS Trichoderma matsutake.  
XX  
PN WO200014242-A1.  
XX  
PD 16-MAR-2000.  
XX  
PF 19-AUG-1999; 99WO-JP04441.  
XX  
PR 08-SEP-1998; 98JP-0270606.  
XX  
PA (NTSB ) JAPAN TORACCO INC.  
PA (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.  
PI Takakura Y, Kuwata S, Ohta S;  
XX  
DR WPI; 2000-256990/22.  
DR N-PSDB; AAA07403.  
XX  
PT Mushroom-derived antibacterial protein against plant pathogenic fungi

PT of rice, with activity and thermal stability, obtainable cheaply on  
PT large scale, useful in agriculture -  
XX  
PS Claim 3; Page 41-45; 52pp; Japanese.

CC This sequence is the Trichoderma matsutake antibacterial protein of  
CC the invention. The protein has activity against at least Pyricularia  
CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an  
CC aqueous extract of a mushroom precipitated by the ammonium sulphate  
CC precipitation method. The protein has a molecular weight of about 210 kD  
CC as determined by the gel filtration method, includes components of about  
CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous  
CC neutral solution at 60 degrees C for 10 minutes but with loss of  
CC antibacterial activity after heating in the solution at 80 degrees C for  
CC 10 minutes. The protein is used for inhibiting the growth of plant  
CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.  
CC It is useful in treating rice plants, and is applicable in agriculture as  
CC an antibacterial agent. The protein has activity at relatively low  
CC concentrations, and can be produced at low cost on large scale.

SQ Sequence 564 AA;

Query Match 33.7%; Score 976.5; DB 21; Length 564;  
Best Local Similarity 41.3%; Pred. No. 2e-83;  
Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;

QY 2 AEEGTAVPVVPGYHKKNETEFKIDRFVNVKLGALQOVSVPRNQNVPTLDPGAWSAAPP 61  
DB 62 aelsgqdnvlgahhrnsikfqtdktdkfvniingalqpsispsdyqptlavaaw-app 120  
QY 62 GSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHWTCTSPRIHPMESLPGIGRPK 117  
DB 121 idpaegqlvimhnpqeaqlnlpgsavtrvgmawhwtcaactphd-----ee 170  
QY 118 LSNDAEDDKENNELYSAEERLIGTSTKEFDESIRHTLVLSQDAYKDRQRIPLPLA 177  
DB 171 rvnnpv-dkqefdalleraktllnvhsdgydsirgqvketlqtl-dasrgvtlplg 228  
QY 178 CHRLNAPYEVVHSAENLFHSIYNDKQKLFLLTNHRCRLALTGGYEKKIGAAEVR 237  
DB 229 verrtdnplyvtwtgad---tvlgdvkpsrfrvltvtrvktfivsetnptqvvaalir 284  
QY 238 NLLATRNPSQLDSYIMAKVYVYVLAIGAIGNPQILYNSGFSGLQVTPRNDSLIPNLGRYIT 297  
DB 285 nl-----ntsn--delvvaqsviaagavctpqilwns-----nirph-----algryls 327  
QY 298 EQPMAPFCQIVLQREFVDSVRDDPYGLPWKEAVAQHIKKNPTDALPIPRDPPEQVTPFP 357  
DB 328 eqsmtfcqivlkrsvdsiatdpr----faakveahkkhpdvlpiphepeqvmipy 383  
QY 358 TTEHPWHTQIHRDAFSYGAAGPEVDSRVIVDLRWFGADPEANLLVF-----QND 408  
DB 384 tsdfpwhvqvr--yafgdvqpkadprvvvdrlfrfgksdiveenrvtfgnpkldweag 441  
QY 409 VODGYSMPQPTFRYRSTASNVARKMMADMCEVASNLGGYLTSPQPMDFGLALHLAG 468  
DB 442 vtdtygmpeptfhvkrtnadgrdqrmdmtnvniigylpgsyppqfmapglaqhitg 501  
QY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGIRTFGFGENPFLTSMCHAIKARSIIINTL 528  
DB 502 ttrigtddqtsvadptskvhnfdnlwvgngcipcdatacnprtstsvayalkgaeavvyl 561  
RESULT 9  
AAW87541  
ID AAW87541 standard; Peptide; 51 AA.  
XX  
AC AAW87541;  
XX  
DT 26-FEB-1999 (first entry)  
XX  
DE Peptide derived from an antitumour protein.





PS Claim 2: Page 27: 52pp: Japanese.

CC This sequence represents an N-terminal fragment of the protein of the

CC invention. The protein is an antibacterial protein with activity against

CC at least *Pycnarularia oryzae* and *Rhizoctonia solani* is obtained from a

CC fraction of an aqueous extract of a mushroom precipitated by the ammonium

CC sulphate precipitation method. The protein has a molecular weight of

CC about 210 kD as determined by the gel filtration method, includes

CC components of about 15 kD and 50 kD in SDS-PAGE, and is stable to

CC heating in an aqueous neutral solution at 60 degrees C. for 10 minutes but

CC with loss of antibacterial activity after heating in the solution at 80

CC degrees C. for 10 minutes. The protein is used for inhibiting the growth

CC of plant pathogenic fungi e.g. *Pycnarularia oryzae* and *Rhizoctonia solani*.

CC It is useful in treating rice plants, and is applicable in agriculture as

CC an antibacterial agent. The protein has activity at relatively low

CC concentrations, and can be produced at low cost on large scale.

xx Sequence 67 AA:

xx

	Query Match	6.1%	Score 177;	DB 21;	Length 67;
	Best Local Similarity	57.4%;	Pred. No.	4.4e-09;	
	Matches	35; Conservative	6;	Mismatches	20; Indels
	Gaps	0;			
Qy	409	VODGYSMPQTFRYRPTSTASNRARKWADMCVEASNIGGYLPTSPPFMDPGLALHLAG	468		
	:	:			
Db	7	vtdcygmppqptfhvkrtnadgdrqrmndntnvanmlggylpgsyqqfmapglvlhitg	66		
Qy	469	T	469		
Db	67	t	67		

RESULT	12
AAB97034	
ID	AAB97034 standard; Peptide: 34 AA.
XX	
XX	
AC	AAB97034;
XX	
XX	
DT	20-JUL-2001 (first entry)
XX	
DE	Lyophilum shimeji antibacterial protein fragment #4.
XX	
XX	
KW	Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
KW	Pycnularia orizae; Rhizoctonia solani; rice pathogen.

XX	Lyophyllum shimeji.
OS	
XX	WG200121657-A1.
PN	
XX	29-MAR-2001.
PD	
XX	20-SEP-2000; 2000WO-JF06404.
PF	
XX	21-SEP-1999; 99JP-0267238.
PR	
XX	(NISB ) JAPAN TOBACCO INC.
PA	(NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
PA	
XX	Takakura Y, Kuwata S, Inoue Y;
PI	
XX	WPI; 2001-281598/29.
DR	
XX	Antibacterial protein and encoded gene isolated from Lyophyllum
PT	shimeji, with activity against plant pathogenic bacteria, applicable in
PT	agriculture e.g. rice cultivation at low concentration, produced at low
PT	cost on large scale -
PT	

Example 2; Page 24; 52pp; Japanese.

CC to ammonium sulphate precipitation. The protein inhibits the growth  
CC of the plant pathogenic bacteria *Pseudomonas* *oryzae* and *Rhizoctonia*  
CC *solani* at a relatively low concentration. *P. oryzae* and *R. solani* are  
CC causative of the two major diseases of rice. The protein contains  
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.  
CC The antibacterial protein can be produced at low cost on a large scale.  
xx  
xx Sequence 34 AA;

```

Query Match      5.8%; Score 168; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. le-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 AERLIGTSTREFDESIRHTLVRLSLQDAYKDRQR 169
      |||||
Db 1 aerligtstrefdesirhtlvrlslqdaykdrqr 34

```

RESULT	13
AAB97031	
ID	AAB97031 standard; Peptide: 30 AA.
XX	
XX	AAB97031;
XX	
XX	
XX	20-JUL-2001 (first entry)
DE	Lyophyllum shimeji antibacterial protein fragment #1.
XX	
XX	
KW	Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
KW	Pyricularia oryzae; Rhizoctonia solani; rice pathogen.

XX	Lyophyllum shimeji.
OS	
XX	
PN	WC200121657-A1.
PN	
XX	
XX	29-MAR-2001.
PD	
PD	
XX	
XX	
PF	20-SEP-2000; 2000WO-JP06404.
XX	
PR	
PR	21-SEP-1999; 99JP-0367238.
XX	
XX	
PA	(NISB ) JAPAN TOBACCO INC.
PA	(NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & F.I.
XX	
PI	Takakura Y, Kuwata S, Inoue Y;
XX	
XX	WPI; 2001-281598/29.
DR	N-PSDB: AAF99980.
DR	

XX Antibacterial protein and encoded gene isolated from *Lyophyllum*  
PT *shimeji*, with activity against plant pathogenic bacteria, applicable in  
PT agriculture e.g. rice cultivation at low concentration, produced at low  
PT cost on large scale -  
XX  
PS Claim 2: Page 32: 52pp; Japanese.

The present sequence is part of an antibacterial protein from the fungus *Lyophyllum shimeji*. The protein was obtained from a fraction prepared by extracting *Lyophyllum shimeji* with water and subjecting the extract to ammonium sulphate precipitation. The protein inhibits the growth of the plant pathogenic bacteria *Pyricularia oryzae* and *Rhizoctonia solani* at a relatively low concentration. *P. oryzae* and *R. solani* are causative of the two major diseases of rice. The protein contains components of 70 kDa and 65 kDa, as determined by SDS-PAGE. The antibacterial protein can be produced at low cost on a large scale.

Query Match	5.6%	Score 162;	DB 22;	Length 30;
Best Local Similarity	100.0%;	Pred. No. 3e-08;		
Matches 30;	Conservative	0;	Mismatches	0;
Indels	0;			





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:51:06 ; Search time 85.12 Seconds  
(without alignments)  
1256.002 Million cell updates/sec

Title: US-09-856-327-2  
Perfect score: 3284  
Sequence: 1 MSLSSTQMLRDPKSMQING.....INTLKGTDGKNTGEHRNL 618

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL19.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phase.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_vertebrate.\*
- 15: sp\_unclassified.\*
- 16: sp\_rvirus.\*
- 17: sp\_bacteriopl.\*
- 17: sp\_cheap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1232.5	37.5	623	3 P79076	P79076 coriolus ve
2	203	6.2	551	2 Q9RH54	Q9RH54 pantoea agg
3	200.5	6.1	573	16 Q9PI90	Q9PI90 campylobact
4	200	6.1	545	2 Q9LBK8	Q9LBK8 gluconobact
5	191.5	5.8	615	2 O34214	O34214 peccobacter
6	190.5	5.8	579	16 Q9A7T6	Q9A7T6 caulobacter
7	182	5.5	591	16 Q9I1K8	Q9I1K8 pseudomonas
8	180	5.5	553	2 Q9XCR0	Q9XCR0 pantoea cit
9	179.5	5.5	529	17 Q9HQ88	Q9HQ88 halobacteri
10	176	5.4	722	16 Q9RZ26	Q9RZ26 deinococcus
11	158.5	4.8	748	10 Q9M0H4	Q9M0H4 arabisdopsis
12	158.5	4.8	748	10 Q9ABP3	Q9ABP3 arabisdopsis
13	155.5	4.7	499	16 Q988P2	Q988P2 rhizobium 1
14	155.5	4.7	769	3 O74253	O74253 pycnopus
15	149	4.5	479	16 Q98D66	Q98D66 rhizobium 1
16	141	4.3	523	16 Q98C76	Q98C76 rhizobium 1

17	138.5	4.2	768	3	O42729	042729 trametes ve
18	127.5	3.9	578	16	Q57307	Q57307 mycobacteri
19	126.5	3.9	772	3	Q12661	Q12661 phanerocoe
20	124.5	3.8	578	16	Q92XY0	Q92XY0 rhizobium m
21	123	3.7	531	16	Q916D0	Q916D0 pseudomonas
22	123	3.7	704	3	Q9P8D9	Q9P8D9 candida tro
23	119	3.6	746	10	Q9LW56	Q9LW56 arabisdopsis
24	118.5	3.6	785	3	Q9P8H5	Q9P8H5 humicola in
25	117.5	3.6	620	3	Q12623	Q12623 humicola gr
26	117	3.6	736	10	Q9ZWB9	Q9ZWB9 arabisdopsis
27	116.5	3.5	698	3	Q9P8D8	Q9P8D8 candida clo
28	116	3.5	502	5	O18672	O18672 caenorhabdi
29	116	3.5	599	5	Q18429	Q18429 caenorhabdi
30	115.5	3.5	502	16	Q92L01	Q92L01 rhizobium m
31	114	3.5	1012	10	O64492	O64492 arabisdopsis
32	113.5	3.5	678	10	O65709	O65709 arabisdopsis
33	112	3.4	520	16	Q988D6	Q988D6 rhizobium l
34	112	3.4	828	3	O74240	O74240 thielavia h
35	112	3.4	1011	2	Q9L5M2	Q9L5M2 salmonella
36	112	3.4	1011	2	Q935H0	Q935H0 salmonella
37	112	3.4	1881	12	Q9DUN3	Q9DUN3 vesicular e
38	111.5	3.4	1899	5	Q9NDY7	Q9NDY7 leishmania
39	110.5	3.4	507	16	Q982L9	Q982L9 rhizobium l
40	110.5	3.4	589	10	Q9M4V5	Q9M4V5 oryza sativ
41	110.5	3.4	5170	5	Q17490	Q17490 caenorhabdi
42	110.5	3.4	6994	5	Q17343	Q17343 caenorhabdi
43	109.5	3.3	1125	5	Q93203	Q93203 caenorhabdi
44	109	3.3	2193	12	Q9QF31	Q9QF31 human coxa
45	108	3.3	1498	16	Q92FP9	Q92FP9 listeria in

ALIGNMENTS

RESULT	1
P79076	
ID	P79076
AC	P79076; PRELIMINARY; PRT; 623 AA.
DT	01-MAY-1997 (TrEMBLrel. 03, Created)
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	PYRANOSE OXIDASE.
OS	Coriolus versicolor.
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC	Aphyllophorales; Coriolaceae; Coriolus.
OX	NCBI_TaxID=57466;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=PS4A;
RX	MEDLINE=971177816; PubMed=9025322;
RA	Nishimura I., Okada K., Koyama Y.;
RT	"Cloning and expression of pyranose oxidase cDNA from Coriolus
RT	versicolor in Escherichia coli.";
RL	J. Biotechnol. 52:11-20(1996).
DR	EMBL; D73369; BAA1119.1; .
DR	InterPro; IPR00205; NAD_binding.
SQ	SEQUENCE 623 AA; 69495 MW; 5D3FC81B35FA5B54 CRC64;

Query Match	37.5%	Score	1232.5;	DB 3;	Length	623;			
Best Local Similarity	45.0%;	Pred. No.	2.7e-91;						
Matches	273;	Conservative	73;	Mismatches	176;	Indels	85;	Gaps	16;
QY	37	DVFIAGSGPIGATYAKLCV	EAGLRVNVNVEIGADSFYAVNAE	EECTAVPYVPGYHKKNEIE	96				
Db	48	DVIVGSGPIGCTYARELV	AGYKVMFIDGIDSGUKI-----	GAHKNTVE	95				
QY	97	FOKIDRFVNVIKALQQVSPVNRQNPV	PLDPCGAWAPPCCSSAISNGKNPHQREFNLS	156					
Db	96	YQKNIDKFVNVIOGLMSVSPVNTLV	ILDSPTSQA--SSFFVRNGSNPEQDPLENLS	153					
QY	157	AEAVTRVGGMSTHTWCTSPRIHPHPP	MESLPGIGRPKPLNDPAE--DKKEWELNLYSEARLI	215					

Db	154	QGVATRVVGGMSTHTWCATPFRDRQ-----RPLLVKDDQDADDAEDWRLYTKAESYF	206
Qy	216	GTSTKEFDESIRHTLVLRSLDAYKDRQRIFRPLPLACHRLKNAPEYVEHWSAENLF--H	273
Db	207	KTGTDQFQKESIRHNLVLNKLAEYKG-QRDFQIPLAATR--RSPTFEVMSANTVFDLQ	263
Qy	274	SIYNDKQKKLFTLLTNHRCVTRLALTGGYEKKIGAAGVARNLLATRNPSQLDSY-----	327
Db	264	NRPTDAPNERFNLFPVACERV-----VRN-----TSNSEIESLHDLI	304
Qy	328	-----INAKVVVLASGAIGNPQILYNSGSL-QVTPRN-DSLIPNLGRYITEQPMAFc	379
Db	305	SGDRFEIKADVFVLTAGAVHQAQLLVNSGFGQLGRDPDPANPQLLPSLGSIYTESLVFC	364
Qy	380	QIVLRQEFVDSVRDDPY--GLP-----WKEAVAQHIKAKNPTDAL	417
Db	365	QTVMSTELIDSVKSDMIIRGNPGDLGYSVYTPGAETNKHDPWNNEKYKNHMQHOEDPL	424
Qy	418	PIPRDPPEQVTTPTTEBHPNHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPANNL	477
Db	425	PIPEDPEQVTTTLFPQSPHPWHTQIHRDAFSYGAVQQSIDSRLLIVDRFFRGTEPKEENK	484
Qy	478	LVFONDVODGYSMQPQTPRYR- PSTASNVARKMADMCVNASNLGGYLPTSPPOFMDPG	536
Db	485	LWFSDKITDTYINMPQPTDFRFPAGRTSKEADMMTDMCVMSAKTGGFLPSLGFPMFEPG	544
Qy	537	LALHLAGTRTGFQDKA--TTVADNNSLVMDFANLVVAGNGTIRTFGFGNPILTSCHAIK	594
Db	545	LVHLGGTHRGFDEQDKCCVNTDSRVFGFNFLGCGGNIPITAYGANPILTAMSLAIK	604
Qy	595	SARSIIN 601	
Db	605	SCVEIKN 611	

RESULT 2

Q9RH54	AC	Q9RH54	PRELIMINARY;	PRT;	551 AA.
ID	ID	Q9RH54;			
DT	01-MAY-2000	(TREMBlrel. 13, Created)			
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)			
DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)			
DE	DEHYDROGENASE SUBUNIT I,				
OS	Pantoea agglomerans.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Pantoea.				
OX	NCBI_TaxID=549;				
RP	[1]				
RN	SEQUENCE FROM N.A.				
RT	Chang K.-I., Yum D.-Y., Pan J.-G., Shin Y.-C.;				
RA	"Cloning and expression of a gene cluster encoding three subunits of				
RT	membrane-bound 2-keto-D-gluconate dehydrogenase from Erwinia herbicola				
RT	ATCC08111 in Escherichia coli.;"				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF068066; AAF21261.1; -				
DR	InterPro; IPR000205; NAD-binding.				
SQ	SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAFBF CRC64;				

Query Match 6.2%; Score 203; DB 2; Length 551;  
 Best Local Similarity 19.9%; Pred. No. 8.7e-08;  
 Matches 132; Conservative 84; Mismatches 250; Indels 198; Gaps 26;

Qy	24	KNATHEYTGNDGVDFIAGSGPIGATYAKLCVEAGLRVVMVEIGAANDSFYAVNAEGTAV	83
Db	2	KKPVFTAGQASADIVIVGSGIVGGMANELVSGQSYLVLEAGL-----	46
Qy	84	PYPVGYHKKKEIEFOKIDRVNVNVIKALQOVSVFVRNQ-----NVPTLDPGAWSAP----	135
Db	47	-----RIDRAQAVENVRN-----HPFANRAGSDFGQLYPSQKFAPAPLYF	86
Qy	136	PGSAINGNKPNHQREFENUSAEAVTRGVGMGSTMHTWCSTPRIHP-----MESLPGIGRPK	192

Db	87	PRNNYV-NVTGPNADSTQ----	QCYLRTVGGTTHHWAASCRHHPSDFVMQSYGYGVR--	133
Qy	193	LSNDPAEDDKEW----	NEL-----YSEARLIGTSTKFEDESIRHTLVLRSLQDAYKDRQRI	245
Db	140	-----DWPIGYDELEPWTYKAENEIGYVAGP--NDPAROSPTERS-----	176	
Qy	246	FRPLPLACHRLKNAPEYVHWSABNLPHSYINDOKKKL-----FTLLLTNH	291	
Db	177	-QPYPM-----DMVPFAHGDNYPASVVPNGHYNLVPIPOGRSTRPWEGRPCTCGNN	226	
Qy	292	RCTRLALTGGYEKKIGAAEVRNLLATRNPSQLDSYIM-----	329	
Db	227	NCQPCIPGAMYNGIHVVE-----RAERNGAVVLAEVYKMDTSDNNRITAVHVLDTSGA	282	
Qy	330	-----AKVYVLASGAIGNPQILYNSGFSGLQVTPRNDSLIPN----	380	
Db	283	SHKATAFAFALACNGIETPRLL-----MAANDANPNGIANASDMVGRNMDHSGFHCS	336	
Qy	381	IVLRQEFVDSVRDDPYGLPWKKEAVAQHIAKNTP-DALPIPRDRPPEQVTPPTTEHPWH	439	
Db	337	FLTKE-----PVM-----LKGGAQSSCMVGYRGGD-----FRDYSAN	370	
Qy	440	TQI-----HRDAFSYGAVGPEVDSRV-----IVDLRFWFGATDPEANLLVFQND	483	
Db	371	KVILNNISRVVTAQQAMKGLVGKALDEIRYRAVHSVDLSISLEPLDPDENRLTFSKT	430	
Qy	484	VODGYSMPQPTFRYPSTASNVRAKRMADCEVASNLGGLYPTSPQFMDPGLAL--HL	541	
Db	431	RKDPHGLPCPDIIYDVGDYVRKGAESAHAQLEHI-----GQLFDAKEFTISQGLNANNHI	485	
Qy	542	ACTTRIGCFDKATTVADNNSLVWDFANLYVAGNGTIRGTGCGENTLTSCHAIKSARSILN	601	
Db	486	MGVIMGKNAKEAVVDGNCRAFDHENLWLPGGGAIPASVYVNSTLTMAALGLKAAHDISL	545	
Qy	602	TLKG 605		
Db	546	RMKG 549		
RESULT	3			
Q9PI90	AD	Q9PI90	PRELIMINARY;	PRT; 573 AA.
IC	99PI90;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	PUTATIVE OXIDOREDUCTASE SUBUNIT.			
GN	CJ0415.			
OS	Campylobacter jejuni.			
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;			
OC	Campylobacter.			
OX	NCBI_TaxID=197;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NCCT 11168;			
RX	MEDLINE=20150912; PubMed=10688204;			
RA	Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,			
RA	Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,			
RA	Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,			
RA	Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,			
RA	Whitehead S., Barrall B.G.;			
RT	"The genome sequence of the food-borne pathogen Campylobacter jejuni			
RT	reveals hypervariable sequences."			
RL	Nature 403:665-668(2000).			
DR	EMBL; ALI39075; CAB74251.1; -.			
KW	Complete proteome.			
SQ	SEQUENCE 573 AA; 63686 MW; 7BEB522A779A5F94 CRC64;			

Query Match	6.1%	Score 200.5;	DB 16;	Length 573;
Best Local Similarity	20.6%;	Pred. No. 1.5e-07;		
Matches 135; Conservative	74;	Mismatches 270;	Indels 175;	Gaps 26;











```
QY 187 GIGRPKLSN-----DPAEDDKENNELYSEARLIGTSTKPEFDESIRHVLVLSLQDA 238
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 294 ---RQRWASEHGLSDVADPGYD---RHIDAVLERM-GVSEQCSHNGPHQ---RLVEGA 342
QY 239 YKDRQIRFRPLPLACHRLKNAPEYVEHSAENLFHSIYNDQ---KOKKLEF-----286
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 343 DKLGYYTFVKA-----ALNLSPEH---YDADKAGHAGFGDGTGAKQGTINTFLKDAFEAG 393
QY 287 --LNTNHRCTRALTGGYERKKGAAEVRNLLATRNPSQLDSYIMAKVYVILASGAIGNPQ 344
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 394 ARILVGTQAQRVLVEDG-----RAAGYSAVTTMGDETRQIT--VRAPOVVVACGALETPA 446
QY 345 ILYNSGSGQLQVTRPNDSLIPNLGRITYEQPMACQIVLRQEFVDSVRDDPYG---LPWW 401
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 447 LLLRSIGGG-----PAAGRYLRLHPAGLVAGI-----YGEDQRAWW 482
QY 402 KEAQAQHIKAKNPTDALPIPRDPE-----PQVTT---PFT--EEH-PWHTOI 442
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 483 GP-----PQSGILKQFADHENGHGFIIEGVQYGPALMASGLPWTGGGAHRDLMSKF 533
QY 443 HRDAFSYGAVGPEVDSRVIVD-----LRWFGATDP-EANLLVFNQNDVQGYSM-----P 491
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 534 HRMATEFSIVQDRGHQGVTVDEGNVAHYALTDDLDARN---FRRGVTESIRLHEAAGA 590
QY 492 QPTERYRPSTASNRARAKMMADMCEVAS---NLGGYLPSPQPMQDGLALHLAGTRIG 548
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 591 EEIIVALAPGVAPWRRGDDLEAFIGQVAVPLGAGGQVFS-----AHQMGSRMG 640
QY 549 FDKATTVADNNSLYWDFANLVVAGNGTIRTGFGENPILTSMCHAIKSARSINTLKGGTD 608
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 641 SDPQTSVADPDQGLHDVPGVWIGDTSAPPTCSGVNPMVSMALASRTAEKLLAAWEGADG 700
QY 609 GKNTG 613
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 701 GTGSS 705

RESULT 11
Q9MOH4 PRELIMINARY; PRT; 748 AA.
AC Q9MOH4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN A74G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL161573; CAB81445.1; -.
DR HSSP; P22637; 3COX.
DR InterPro; IPR000205; NAD_binding.
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81959 MW; 13CD183F5940DAC3 CRC64;

Query Match 4.88; Score 158.5; DB 10; Length 748;
Best Local Similarity 19.48; Pred. No. 0.00058;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFTAGSGPTGATYAKLCVEAGLRVWVEIGAADSFTAVNAEETATVPYVPGYHKHKEIE 96
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
Db 239 DAVVVGSGGGVAAANAKAGLVLEKG---NYFTHADYSGLEVPMSMLEYKKG---292
QY 97 FQKDIDRFVNVYIKGALQOVSVFVRNQNVPTLDPGAWSPGSSAISNGKNPHQREFENLS 156
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 293 -----GLLITV-----DCK-----FMLLA 306
QY 157 AEAVTRGVGGMSTHWCTSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSEARLIG 216
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----QEW---SEGSKIF 340
QY 217 TSTKEF---DE-SIRHTLVLSLQDAYKDKQRIFRPLPLACHRL-----KNAP--261
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 341 FGSQEQSAMDVEIRIGVTGTERCVKHGFQ---QVLRKGCERLGLQVESVPRNSPEDH 395
QY 262 -----YVEHSAENLFHSIYNDQKQKLLFTLLNHRCTRLLAL-----TGGYEKK---I 306
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 396 YCGLGCGYGRAGAKNGTDQTLWLDVAVENGAVILTGKAEFRVLVDNNTSSNERKKRCVGV 455
QY 307 GAAEVRNLLATRNPSQLDSYIMAKVYVILASGAIGNPOILYNSGSGQLQVTRPNDSLIPN 366
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 456 PASSVGKGKIGK-----FIIEARVTVSSAGSLLTPLPLMLSSGLKN-----PN 497
QY 367 LGR-----YITEQPMACQIVLRQEFVDSVR--DDPYGLPWWKEAQAQHIKAKN 412
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 498 IGRNLKHPVLMTWGYPEKSEFSGKMYEGGIITSVHHMND-----TESGCKAILLEN 550
QY 413 PTDALPIPRDPEPQVTPPTTEEHPHWTQIHRDAFSGAVGPEVDSRVIVLRFWFGADP 472
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 551 PLIG-PASYAGLSPWVS-----GPDLEKRIK-----YGR--580
QY 473 EANNLLVFQNDVQGYSM--PQPTFRYRPSTASNRARAKMMADMCEVAS---NLGGY---524
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 581 --AHLFALVRLDLSGGEVNMENEVYRTTKDRENLRAGLRVSAAGAAGEVGTGRSD 638
QY 525 -----LPTSPQEMDPGLAL-----HLAGTTRIGDFDKATTVADNN 559
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 639 GQKMKCEAITKEAMEEFLEVDVAVGCVGTGKEGYWTFYSAHOMGSCRMGVTAEGALDEN 698
QY 560 SLVWDFANLYVAGNGTIRTGFGENPILTSMCHAIKSARSINTLKGGT 607
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 699 GESWEAEGLFVCDGSLPSAVGVNPMITIQSTAYCISISKIVDSLQNK 746

RESULT 12
Q94BP3 PRELIMINARY; PRT; 748 AA.
AC Q94BP3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN A74G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene A74g28570 (GI:7269712).";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY039977; AAK64154.1; -.
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81987 MW; E3EB613F5D56A0D1 CRC64;
```

```
Query Match      4.8%; Score 158.5; DB 10; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00058;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEIE 96
Db 239 DAVVVGSGGGVAAANLAKAGLVLEKQ---NYFTHADYSGLEVPSMLELYEKG--- 292
QY 97 FOKDIDRFVNVKIGALQQVSVPRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156
Db 293 -----GLLTVV-----DGK-----FMLLA 306
QY 157 AEAVTRGVGMSTHWTCTSPRIHPMESLPGIGRKLNSNDPAEDDKWENELYSAEERLIG 216
Db 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----QEW-----SEGSKIRF 340
QY 217 TSKTEF---DE-SIRHTLVLSQDAYKDRQIRFRPLPLACHRL-----KNAPK--- 261
Db 341 FGSQEQYQSAMDVETIRIGVTERCVKHGFQ---QVLRKGCERGLQLOVESVPRNSPEDH 395
QY 262 -----YVHWSAENLFHSIYNDKQKLFLLTHNRCTRRLAL-----TGGYEKK---I 306
Db 396 YGCLGCGYGRAGAKNGTDQTLWDVAVENGAVILTGKAERFVLVDMTSSNKKRKCQGV 455
QY 307 GAAEVNLLATRNPSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIPN 366
Db 456 FASSVGGKIGKK-----FIIEARVTVSSAGSLTTPPLMLSSGLKN-----PN 497
QY 367 LGR-----YITEQPMAFQIVLRQEFVDSVR---DDPYGLPWWKEAVAQHIKAKN 412
Db 498 IGRNLKHLVPLMTWGYFPEKDEFSEFGKMYEGGIITSVHMD-----TESGCKAILEN 550
QY 413 PTDALPIPPRDEPOVTPTEEHHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWCGATDP 472
Db 551 PLIG-PASTAGLSPPWVS-----GPDLMKRMK-----YGR--- 580
QY 473 EANNLLVFQDQGYSM--POPTFRYRPSTASNVRAKMDMCEVAS---NLGGY--- 524
Db 581 --AHLFALVRLDLSGVMNEVTVTKKDRNLRAGLRQALRVSAAGAVEVGYRSD 638
QY 525 -----LPTSPQFMDPGAL-----HLATTRIGFDKATTVADNN 559
Db 639 GQMKCEAITKAMEEFLDEVDAVGSGVTKGEYTWYTFSAHQMGSCRMGVTAEEGALDEN 698
QY 560 SLVWDFANLYVAGNGTIRGFGFNPTLSTSMCHAIRKARSIIINTLKGTT 607
Db 699 GESWEAEGLFVCDGSLPSAVGVNPMITIQSTAYCISSKIVDSLQNK 746

RESULT 13
Q988P2
ID Q988P2 PRELIMINARY; PRT; 499 AA.
AC Q988P2;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DE MLC6655 PROTEIN.
GN MLC6655.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
```

```
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003009; BAB52905.1; -.
DR InterPro; IPR00205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 499 AA; 54537 MW; 633924AED96B8F51 CRC64;

Query Match      4.7%; Score 155.5; DB 16; Length 499;
Best Local Similarity 20.7%; Pred. No. 0.00053;
Matches 129; Conservative 65; Mismatches 235; Indels 195; Gaps 27;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEIE 96
Db 7 DIVIIGSGSSLAAYSLADTGRIIVILE-----RGEHLRDTPE 45
QY 97 FOKDIDRFVNVKIGALQQVSVPRNQNVPTLDPGAWSPGSSAIS-----SNCK--- 145
Db 46 ARDDIAIFQNGFYRSEE-----WLATDGESFLPGNTYYVGGNSKFFG 88
QY 146 -----NPHOREFENLSAEAVTRGVGMSTHWTCTSPRIHP---PMESLPGIGRP 191
Db 89 AVMYRYRQEDFNPRDH-----MGRRSPQNPISYAELEPWYERAELLFGV-RG 134
QY 192 KLSNDPAEDDKWENELYSAEERLICTSTKEFDESIRHTLVLSQDAYKDRQIRFRPLPL 251
Db 135 DARQDPTPEPR--NRPY-----RYLPVPDEPAITATVQRLLQAGIHFA-----SLPL 179
QY 252 A-----CHRLKNAPEYVHWSAENLFHSIYNDKQKLFLLTHNRCTRRLALTGGYEKK 305
Db 180 AIDDAWLRRAKGT-----WDAFPNT--GAGKIDAEVGLTKALEHPNATLITGANVORL 232
QY 306 IGAAEVNLLATRNPSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIP 365
Db 233 VTDSGRRVMAAVFKDGVLSIGADVFAVAGAVQSAALLRS-----STSVTP 282
QY 366 N-----LGR-YITEQPMAFQI-----VLRQE---FVDSVRDDPYG-LPMWKEAV 405
Db 283 NGLGNSSDQGRNFMNHNNTAMLAIDPFRNRTAVYQKTLGFNDYFNKQDPLGSPFLGNVOL 342
QY 406 AQHIKAKNPDA---LPIPRDEPOVTPTEEHHPWHT-QIHRDAFSYGAVG---PEVD 457
Db 343 LHGITGNTILKANAPLLP-----RWLAGLVARNCYGVWFLTSEDLPNPE 384
QY 458 SRVIVDLRFWCGATDPDPAENLLVFQDQGYSMPOPTFRYRPSTASNVRAKMM--ADMC 515
Db 385 SRVTI-----RNGRIVMNVVRNMCAHETLIR-----RTRAVMREAGFP 423
QY 516 EVASNLGGYLPTSPQFMDPGALHLAGTRIGFDKATTVADNNSLVWDFANLYVAGNGT 575
Db 424 VVLTTRTEGRKTTVS-----HQCGRALGSDPNTSVSPDCRSHDIANLYVTDAV 472
QY 576 IRTGFGFNPTLSTSMCHAIRKARSII 599
Db 473 LPTSAVNPALTIATAALAKAGAAI 496

RESULT 14
O74253
ID O74253 PRELIMINARY; PRT; 769 AA.
AC O74253;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE CELLOBIOSE DEHYDROGENASE (EC 1.1.99.18).
GN CDH.
OS Pycnoporus cinnabarinus.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Stereales; Schizophyllaceae; Pycnoporus.
OX NCBI_TaxID=5643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99321795; PubMed=10393235;
```



Db 429 VVDANCKVHGMNLYVGGSSVFCTSGQANPTTT 461

Search completed: September 3, 2002, 16:00:34  
Job time: 568 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:55:51 ; Search time 28.03 Seconds  
(without alignments)  
853.681 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLTEQMLRDYPRSMQING.....IINTLKGGTDGKNTGHEHRL 618

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126.5	3.9	773	1	CDH_PHACH
2	116	3.5	514	1	G6PD_MYCTU
3	107.5	3.3	964	1	XINO_YEAST
4	103.5	3.2	612	1	AMTG_ASPOR
5	103.5	3.2	639	1	AMTG_ASPAK
6	102.5	3.1	1536	1	SIN3_YEAST
7	101	3.1	1010	1	SCA4_RICPA
8	100	3.0	663	1	ALOX_CANBO
9	100	3.0	1142	1	ENRM_PIG
10	98.5	3.0	382	1	MTLD_KLEPN
11	98.5	3.0	556	1	BETA_ECOLI
12	98	3.0	505	1	Y4NJ_RHISN
13	98	3.0	925	1	W70T_HUMAN
14	98	3.0	1257	1	CCAA_BACTU
15	96.5	2.9	684	1	Y492_MYCTU
16	96	2.9	291	1	US02_HSV11
17	95.5	2.9	382	1	MTLD_ECOLI
18	95.5	2.9	454	1	DLHD_RHOCA
19	95.5	2.9	519	1	AMVH_SACFI
20	95.5	2.9	708	1	HELS_SULSO
21	95.5	2.9	883	1	RPOL_BPT7
22	95	2.9	487	1	ENGA_CHLPN
23	95	2.9	500	1	YDAK_YEAST
24	94.5	2.9	855	1	BUN2_DROME
25	94.5	2.9	1211	1	GAF1_SCHPO
26	94.5	2.9	2193	1	POLG_CXAL6
27	94	2.9	606	1	PRIM_MYXXA
28	94	2.9	886	1	SMGB_MOUSE
29	93.5	2.8	406	1	YNQ5_YEAST
30	93.5	2.8	594	1	CIK1_YEAST
31	93.5	2.8	823	1	SCH9_YEAST
32	93.5	2.8	903	1	VGLB_HSV1F
33	93.5	2.8	904	1	VGLB_HSV11

34	93.5	2.8	904	1	VGLB_HSV1P
35	93.5	2.8	1859	1	RPB1_CAEEL
36	93	2.8	470	1	LEU2_AZOVI
37	93	2.8	552	1	CHOD_BREST
38	93	2.8	605	1	GOX_ASPNG
39	93	2.8	737	1	AMY1_AEDAE
40	93	2.8	890	1	GLND_ECOLI
41	93	2.8	890	1	GLND_SALTY
42	93	2.8	953	1	YNM7_YEAST
43	93	2.8	955	1	VP2_BTIV17
44	93	2.8	1012	1	UBAL_SCHPO
45	93	2.8	1117	1	CYT4_NEUCR

#### ALIGNMENTS

```
RESULT 1
CDH_PHACH STANDARD; PRT; 773 AA.
AC Q01738; O00047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cellobiose dehydrogenase precursor (EC 1.1.5.1) (CDH) (cellobiose-
DE quinone oxidoreductase).
GN CDH-1 AND CDH-2.
OS Phanerochaete chrysosporium.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Corticiaceae; Phanerochaete.
OX NCBI_TaxID=5306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OGC101;
RX MEDLINE=97077226; PubMed=8919793;
RA Li B., Nagalla S.R., Renganathan V.;
RT "Cloning of a cDNA encoding cellobiose dehydrogenase, a
RT hemoflavoenzyme from Phanerochaete chrysosporium.";
RL Appl. Environ. Microbiol. 62:1329-1335(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OGC101;
RX MEDLINE=97176414; PubMed=9023960;
RA Li B., Nagalla S.R., Renganathan V.;
RT "Cellobiose dehydrogenase from Phanerochaete chrysosporium is encoded
RT by two allelic variants.";
RL Appl. Environ. Microbiol. 63:796-799(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 19-208.
RX MEDLINE=20139694; PubMed=10673428;
RA Hallberg B.M., Bergfors T., Boeckbro K., Pettersson G., Henriksson G.,
RA Divine C.;
RT "A new scaffold for binding haem in the cytochrome domain of the
RT extracellular flavocytochrome cellobiose dehydrogenase.";
RL Structure 8:79-88(2000).
CC -|- FUNCTION: DEGRADATES BOTH LIGNIN AND CELLULOSE. OXIDIZES CELLOBIOSE
CC TO CELLOBIONOLACTONE.
CC -|- CATALYTIC ACTIVITY: Cellobiose + a quinone = cellobiono-1,5-
CC lactone + a phenol.
CC -|- COFACTOR: ONE FAD AND ONE HEME B.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMC
CC OXIDOREDUCTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U46081; AAC49277.1; -
```

727 VKVFGTNNLFIYDAGIIPHLPTG---NPQGTILMSAAEQAAAKIL-ALAGG 772

```

RESULT 2
G6PD_MYCTU
ID G6PD_MYCTU STANDARD; PRT; 514 AA.
AC 008407;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
DS ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCY493.07.
GN Mycobacterium tuberculosis
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
SEQUENCE FROM N.A.
RP RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsley T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
SEQUENCE FROM N.A.
RP RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -1- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE
CC A CLASSICAL ZWF.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; Z95844; CAB09259.1; -.
CC EMBL; AE007019; AAK45757.1; -.
CC HSSP; P11411; 1DPG.
CC TIGR; MT1494; -.
CC TubercuList; RV1447c; -.
CC InterPro; IPR001282; G6PD.
CC Pfam; PF00479; G6PD; 1.
CC Pfam; PF02781; G6PD_C; 1.
CC PRINTS; PR00079; G6PDHGRGNASE.
CC PRODOM; PD001129; G6PD; 1.
CC PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Glucose metabolism; Complete proteome.
FT ACT_SITE 210 BY SIMILARITY.
SQ SEQUENCE 514 AA; 57343 MW; 22A9CCEDC9AB062F CRC64;

```



CC	EMBL; Z38059; CAA86148.1; -	DR	PIR; S48404; S48404.	DR	HSSP; P08657; 1CLD.	DR	SGD; S0001392; YII130W.	DR	InterPro; IPR001138; Zn2_Cy6_fungal.	DR	Pfam; PF001172; Zn_c1us; 1.	DR	PRINTS; PR00054; FUNGALZNCYS.	DR	SMART; SM00066; GAL4; 1.	DR	PROSITE; PS00463; Zn2_Cy6_FUNGAL_1; 1.	DR	PROSITE; PS50048; Zn2_Cy6_FUNGAL_2; 1.	KW	Hypothetical protein; Transcription regulation; DNA-binding;	KW	Nuclear protein; Zinc; Metal-binding.	FT	DOMAIN 21 47	FT	DOMAIN 811 896	FT	SEQUENCE 964 AA; 108780 MW; AD5ABE59B4B022CC CRC64;	QY	33	NDGVDFVFIAGSGPIGAYAK--LCVEAGLRVVMVEIGAADSFYAVNAEECTAVPYV----	DB	322	DEGYKFIARKLIDITNARDLSNIAQLMLII-----FLOCSARLSTCTVYIGVAM 373	QY	87	-----PGYHKK-----NEIEFOK-----DIDRFVNVIKGALQQVSVPRNQNPV 125	DB	374	RSALRAGFHRKLSPNSGFSPIEIMRKRLFTYYIKLDVINAMGLPRISIPDDPDTLP 433	QY	126	TLD-----PGWASAPPSSAISNGKNPHOREF-----ENLSAEAVTRGVGMST 169	DB	434	-LDLSDENITEVAYLPENQHSLSSTGIS--NEHTKLFILNEIISELYPIKKTNNIIS 489	QY	170	HWTCT--PRTHPMESLPGIGRPKLSN-DPAEDDKENNELYSEARLGTSTKEFDESI 226	DB	490	HETVTSLELRKLNWLDLKLPIPNAENIDPE-----YERANRLHL-----SF 533	QY	227	RHTLVRLSLQDAYKDRQIRFPL-----PLACHRLKN-----APE 261	DB	534	LHVQII-----LYRPFHYLSRNNAENVDPICRYRARSIAVARTVIKLAKE 581	QY	262	YVE-----WHSANLFTSIY-----NDDKQKLFLLTLNHRCTRALTGG 301	DB	582	MVSNLLTGSVWYACYTFYSVAGLLFYIHEAQLPKDSAREYVDILKDAETGRSVLIQL 641	QY	302	YEKKIGAERYNLLATRNPSQLDSYIMAKVYVILASGAINPOLYN--SGFSGLOVTPR 359	DB	642	KDSSMAASRTYNLL-----NQIFEKLNSTKIQLTA-----LHSSPSNESAFVLTNN 687	QY	360	NDSLIPNLGRVITBOPMAFCQIVLRQEF-----VDSVRDDPYGLPWKKEAVAOHIAKNPT 414	DB	688	SSALKPHLGDSL-QPPVFSSQDTKNSFLAKSEESTND-----YAMANYLNTPI 737	QY	415	DALPIPRDPPEQV---TPTFTEHPWHQTIHRDAFSYGAGVPEVDSRVIVDLRWFGATD 471	DB	738	SENPLNEAAQQDQVSGQGTNNMSNE-----RDPNFLSIDIRLDNN----- 777	QY	472	PEANLLVFQNDV---ODGYSMPQPT---FPRYRSTASN 504	DB	778	-GOSNILDATDDVFIRNDG-DIPTNSAFDFSSSKSNASN 814	RESULT 4	AMYG_ASPOR	ID	AMYG_ASPOR	STANDARD;	PRT;	612 AA.
AC	P36914;	DT	01-JUN-1994 (Rel. 29, Created)	DT	01-JUN-1994 (Rel. 29, Last sequence update)	DT	30-MAY-2000 (Rel. 39, Last annotation update)	DE	Glucosylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)	DE	1,4-alpha-D-glucan glucohydrolase).	GN	GLAA.																																																																																			



between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D00427; BAA00331.1; --  
PIR; J00479; J00479.  
HSP; P04064; IGAI.  
GlycoSuiteDB; P23176; --  
InterPro; IPR002044; CBD.4.  
InterPro; IPR000165; Glyco\_hydro\_15.  
Pfam; PF00686; CBD\_4; 1.  
Pfam; PF00723; Glyco\_hydro\_15; 1.  
PRINTS; PR00736; GLYDRLASE15.  
ProDom; PD001568; CBD\_4; 1.  
ProSITE; PS00820; GLUCOAMYLASE; 1.  
Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
Signal.  
KW SIGNAL 1 18  
FT PROPEP 19 24  
FT CHAIN 25 639  
FT DOMAIN 494 538  
FT BINDING 143 143  
FT ACT\_SITE 199 199  
FT ACT\_SITE 202 202  
FT ACT\_SITE 203 203  
FT DISULFID 233 236  
FT DISULFID 245 472  
FT DISULFID 285 293  
FT DISULFID 194 194  
FT CARBOHYD 418 418  
FT CARBOHYD 464 464  
FT CARBOHYD 466 466  
FT CARBOHYD 467 467  
FT CARBOHYD 475 475  
FT CARBOHYD 476 476  
FT CARBOHYD 482 482  
FT CARBOHYD 483 483  
FT CARBOHYD 485 485  
FT CARBOHYD 487 487  
FT CARBOHYD 491 491  
FT CARBOHYD 495 495  
FT CARBOHYD 498 498  
FT CARBOHYD 499 499  
FT CARBOHYD 500 500  
FT CARBOHYD 501 501  
FT CARBOHYD 503 503  
FT CARBOHYD 505 505  
FT CARBOHYD 507 507  
FT CARBOHYD 511 511  
FT CARBOHYD 512 512  
FT CARBOHYD 513 513  
FT CARBOHYD 514 514  
FT CARBOHYD 516 516  
FT CARBOHYD 517 517  
FT CARBOHYD 519 519  
FT CARBOHYD 521 521  
FT CARBOHYD 523 523  
FT CARBOHYD 524 524  
FT CARBOHYD 525 525  
FT CARBOHYD 526 526  
FT CARBOHYD 527 527  
FT CARBOHYD 528 528  
FT CARBOHYD 529 529  
FT CARBOHYD 530 530  
FT CARBOHYD 531 531  
FT CARBOHYD 533 533  
FT CARBOHYD 534 534  
KW SIGNAL 639 AA; 68271 MW; E112B31A4DD8DD6B CRC64;  
SEQUENCE

Query Match 3.2%; Score 103.5; DB 1; Length 639;  
Best Local Similarity 18.1%; Pred. No. 4.2;  
Matches 119; Conservative 68; Mismatches 233; Indels 239; Gaps 28;  
QY 22 IPKNAIHEITGNDGVDFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFAVNAEET 81  
DB 37 VARTAILNNTIGADG--AWVSGA-----DSGIVASPTDNDPDYFTYTRSGSL 82  
QY 82 AVPYVPGYHKKKEIEFQKIDREVNVIKALQVQVSPVRNQNVPVLPDGAWSAPGSSAI 141  
DB 83 VIKTLVDLFRNGDPTDLLSTIEHYI-----SSGIGEPEKFNVDYETAYT 140  
QY 142 SNKNPHQREFEENLSAEAVTRGVCGMSTHTWCSTPRHPMESLPGIGRPKLSDNPAEDD 201  
DB 107 -----SSQAIIGVSNPSGDL-----SSGIGEPEKFNVDYETAYT 140  
QY 202 KENNELYSEAEELIGTSTKEFDESIRHTLVLSLQDAKYD--RQIFRPLPLACHRLKNAP 260  
DB 141 GSGRPQPDGPALRATAMIGFQVL-----LDNGYTSAAETIVWPL-----VRNDL 186  
QY 261 EYVE--WHSAEENLPHSYNDKQKFLTLNHRCTRALTGG--YEKKIGAAEVRNLLA 316  
DB 187 SYVAQYWNQGT---YDLWEENVSGSFFTIAVOHR---ALVEGSAFATAVGSS----- 232  
QY 317 TRNPSSQLDSYIMAKVYVLASCAIGNQIL--YNSGFSGLQVTPRNDSLIPN-----LG 368  
DB 233 ----CSWCDQA-----FQILCYLQSFWTGSIYLANFDSRRSGKDTNTLLG 274  
QY 369 RYIEQPMAFQ-----IVLRQEFVDSYR-----DDPY 396  
DB 275 SIHTFDEACGDDSTFQPCSPRALNKHVVDSFRSYTLNDGLSDSEAVAGRPEDSY 334  
QY 397 --GLPWPKEAQAQIAKNPTDPIPRPEPOVTPFTTEHPWHTQ-----IH 443  
DB 335 YNGNPWFQSTLA--AEQLYDAL-----YQDKGSLDITDVSIDFF 374  
QY 444 RDAFSYGAVG-----PEVDSRVIVDLRWFG-----ATDPEANLLVFQNDVQDG--YS 489  
DB 375 KALYSGAATGTYSSTSYSSIVSAVKTFADGFGYSIVETHAASNGSLSEQDKSDGDEL 434  
QY 490 MPQPTFRPSTASNVARKMM-----ADCEVASNLGGY---LPTSPQPMQ 534  
DB 435 ARDLTWSYAALLTANNRNSVVPFSGWETSSASWPGTCAATSGTSSVTVTSWPSIVA 494  
QY 535 PGLALHLAGTTRIG----FDKATTVADNNSLVWDFAN-----LYVAGNGTIRTFGFGN 583  
DB 495 TGGTTTATTGGGVTSTKSTTTTASCTSTTSSTSTCTTPTTAVATFDLTATTYGEN 553  
RESULT 6  
SIN3\_YEAST  
ID SIN3\_YEAST STANDARD; PRT; 1536 AA.  
AC P22579; Q08049;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Paired amphipathic helix protein.  
GN SIN3 OR SDI1 OR UME4 OR RPD1 OR GAM2 OR SDS16 OR YOL004W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RC STRAIN=GRF88;  
RP MEDLINE=91042523; PubMed=2233725;  
RX Wang H., Clark I., Nicholson P.R., Herskowitz I., Stillman D.J.;  
RT "The Saccharomyces cerevisiae SIN3 gene, a negative regulator of HO,  
RL contains four paired amphipathic helix motifs.";  
RN Mol. Cell. Biol. 10:5927-5936(1990).  
RP [2]  
RA SEQUENCE FROM N.A.  
RA Hughes B., Pohl T.M.;

Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND  
 CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STE6,  
 CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SPO13), GENES  
 CC THAT RESPOND TO EXTERNAL SIGNALS (PHO5) AND TRK2. IT IS PROBABLY  
 CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR  
 CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3.  
 CC - SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL  
 CC REGULATORY PROTEIN RPD3.  
 CC - SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.  
 CC - DOMAIN: CONTAINS 4 PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A  
 CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A  
 CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.  
 CC - SIMILARITY: TO S.POMBE SPAC12C2.10C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M36822; AAA34839.1; -  
 DR EMBL; 274746; CAA99003.1; -  
 DR PIR; S12068; RGBYS3.  
 DR SGD; S0005364; SIN3.  
 DR InterPro; IPR003822; PAH.  
 DR Pfam; PF02671; PAH; 3.  
 KW Repeat; Transcription regulation; Repressor; Cell division;  
 KW Activator; Nuclear protein.  
 FT DOMAIN 239 1200 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.  
 FT REPEAT 239 252 HELIX A.  
 FT REPEAT 274 285 HELIX B.  
 FT REPEAT 426 439 HELIX A.  
 FT REPEAT 461 472 HELIX B.  
 FT REPEAT 679 692 HELIX A.  
 FT REPEAT 714 725 HELIX B.  
 FT REPEAT 1152 1165 HELIX A.  
 FT REPEAT 1189 1200 HELIX B.  
 FT DOMAIN 480 519 GLN-RICH.  
 FT CONFLICT 510 510 Q -> QAO (IN REF. 2).  
 SQ SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;  
 Query Match 3.1%; Score 102.5; DB 1; Length 1536;  
 Best Local Similarity 19.3%; Pred. No. 19;  
 Matches 115; Conservative 83; Mismatches 192; Indels 207; Gaps 31;  
 QY 72 FYAVNAEGTAVYVPGYHKKNEIEFQKDIDRFVNV---IKGALQOVSVPRNQNVPTLD 128  
 DB 985 FYDILCLADFTFHTTAYSNPKERLKLKLVFISLFFSISPEKTESLYSHKQNV--- 1040  
 QY 129 PGAWSAPGS---SALSNGKNPHOREFENLSA-----EAVTRGVGGMSTHTCTSPRTH 179  
 DB 1041 ----SSSGSDGSSTASRRYQDQEMSLDILHRSRYQKLKR-----SNDEGKVPQLS 1091  
 QY 180 PMWESLPG-IGRPKLSNDPAEDD-----KEM 204  
 DB 1092 EPPEEPNTIEEELIDEAKNPWLGNLVEANSOGIIONRSIFNLFANTNIYFFRW 1151  
 QY 205 NELYSEARLGTSTKEFDESIRHTLVLSRL-----QDAYKD 241  
 DB 1152 TWIY---ERLL--ETQOMNERTVKEINTRSTVTFAKDLDLLSSQLSEMGLDFVGDAYKQ 1206  
 QY 242 RORIFRPLPLACHRLKNAPEYVWEHSAENLFHSIYNDKQKLFIL-----LTHRCR 295  
 DB 1207 VLRLSR-----RLNGDLEQWF--EESLROAVN--NKAFLYIDKVTQSLVKAHPT- 1255  
 QY 296 LALTGGYEKKGAEEVRNLLATRNPSQLDSYIMAKVYVVLASGATGNPOILYNSGFSGLQ 355  
 DB 1256 -LMTDAKTAIMEALFVKD-----RN-----ASTSAKOQIIIR-----LQ 1289

QY 356 VTFRNDSLIPNLGR-----YITEQPMACQIVLROEFVDSVRDDPYGLPWKEAVQA 407  
 DB 1290 VR-SHMSNTENNFRFEDKRTLRHSIQYIALDDLTLKPEKADEK-----WKYYVTS 1340  
 QY 408 HTAKNPTDALPTFPDRPEQVTPTEEPHWHTQIHRDAFSYGAVGCPEDSVRIVDLRWF 467  
 DB 1341 YALPHPTGEG-----PHEKLKIPFE-----RLIEFGQDIDGTEYDEF----- 1379  
 QY 468 GATDPEANNL-----LVFQNDVQDG-YSM--PQPTFRYRPTASTNVRKAMWDMC 515  
 DB 1380 ---SPEGISVSLTKIKIPIYQLHIENGSDYVFRKATNKY-PTIANDNTOKGHVSQKK 1435  
 QY 516 EVASNLGGYLPSPQFMD--PGLALHLAGTTRIGFDK-----ATTVDNNSL 561  
 DB 1436 ELIS-----KFLDCAVGLRNLDLDEAQLSMQKKWENKLSIAKTSAGNQGI 1481  
 RESULT 7  
 SCA4\_RICPA  
 ID SCA4\_RICPA STANDARD; PRT: 1010 AA.  
 AC Q9AJ75;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)  
 DE (Protein PS 120) (Fragment).  
 GN SCA4 OR D.  
 OS Rickettsia parkeri.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=35792;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sekeyova Z., Roux V., Raoult D.;  
 RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the  
 RT 'gene D' coding for an intracytoplasmic protein.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF155059; AAK30690.1; -  
 KW Antigen.  
 FT NON\_TER 1 1  
 FT NON\_TER 1010 1010  
 SQ SEQUENCE 1010 AA; 110694 MW; 78586D8C92FF9C5B CRC64;  
 Query Match 3.1%; Score 101; DB 1; Length 1010;  
 Best Local Similarity 18.2%; Pred. No. 13;  
 Matches 106; Conservative 88; Mismatches 243; Indels 144; Gaps 25;  
 QY 136 PGSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHTCTSPR---IHPPMESLPGIGRP 191  
 DB 55 PMSVLSNGISPSQ-----TSDPITKAV-----RETIQPKQNLIEQILDLAAL--- 100  
 QY 192 KLSNDPAEDDKENNELYSEARLIGT---STKGF-DESIRHTLVLSLQD-----AYKD 241  
 DB 101 -TDRDLAEOKRKEIEEKEKDKTLSTFFCNPANRREIDKALENPELKKLESIEIAGYKN 159  
 QY 242 RORIFRPLPLACHRLKNAPEYVWEH---SAENLFHSIYNDKQKLFIL----- 287  
 DB 160 VHNTES---AASGYPGGFKPVQWENHVSASDLRATVVKNDAGDELCTLNETTVTKPPT 215  
 QY 288 LTNHRCRTRALTGGYE-----KKIGAEVRNLLATR---NPSQLDSYIMAKVYVL 335  
 DB 216 LAKQDGTQVQISSYREIDFPPIKLDKADGSMHL-SMVALKADGTKPSKDPVYFTAH---Y 271



RL Adv. Dent. Res. 10:111-118(1996).  
 CC -!- FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS  
 CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE  
 CC COVERING VERTEBRATE TEETH.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.  
 CC INTACT ENAMELIN AND LARGE-MOLECULAR-WEIGHT ENAMELINS ARE LIMITED  
 CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,  
 CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER  
 CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CRYSTALLITES IN ROD  
 CC AND INTERROD ENAMELIN.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE  
 CC TRANSITION STAGE.  
 CC -!- PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL  
 CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN  
 CC OCCURS SOON AFTER SECRETION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U52196; AAD10837.1; -  
 DR GlycoSuiteDB; O97939; -  
 KW Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.  
 FT SIGNAL 1 38  
 FT CHAIN 39 1142 ENAMELIN  
 FT CHAIN 39 ? 56 KDA ENAMELIN.  
 FT CHAIN 39 ? 89 KDA ENAMELIN.  
 FT CHAIN 39 ? 142 KDA ENAMELIN.  
 FT CHAIN 39 ? 155 KDA ENAMELIN.  
 FT CHAIN 174 276 32 KDA ENAMELIN.  
 FT CHAIN 515 665 25 KDA ENAMELIN.  
 FT CHAIN 670 ? 34 KDA ENAMELIN.  
 FT CHAIN ? ? 45 KDA ENAMELIN.  
 FT MOD\_RES 53 53 PHOSPHORYLATION (PROBABLE).  
 FT MOD\_RES 191 191 PHOSPHORYLATION.  
 FT MOD\_RES 216 216 PHOSPHORYLATION.  
 FT MOD\_RES 547 547 HYDROXYLATION.  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 680 680 H -> D (IN REF. 2).  
 FT CONFLICT 838 840 RDH -> TTI (IN REF. 2).  
 SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;  
 -----  
 Query Match 3.0%; Score 100; DB 1; Length 1142;  
 Best Local Similarity 19.2%; Pred. No. 18;  
 Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;  
 QY 76 NAEETGAVPYVP--GVH-----KKNELEFQKIDTR-----FVNVIKG 110  
 DB 192 NEEGNN--PYGFFGHHGEGGPPYYSEMFQDEKPEKDPKTPATPSVN----245  
 QY 111 ALQGVSVVRNVPTL-DP-GAWSAPGSSAISGNKPNQR-----EFENLSAEAVTR 162  
 DB 246 ----TTVPETNTQPNAPNPRGNDTSPGTGS--GQPNPRSNPTGQNGPANNVSGGVPR 299  
 QY 163 -----GVGGMSTHWTCSPRI-----HPPMELSLPGIGR-----PKLSNDPAEDDKW 204  
 DB 300 SOSPPGPRQTIITHENYPNIRGFFPARRQWRPPG----PAMGHRHNGPFYRNQOIQRGPPW 356  
 QY 205 NELYSEAERLICTSKEFDESIRHTLVLRSLQDAYKDRORIFRPLPLACHRLKNAPEYVE 264  
 DB 357 NSFTLEGG-----QAVRPGIPTTRVY-----GSTARSNPNYA- 390

QY 265 WHSAENLFHSIYNDKQKKLFTLLNNHRCTRALTG-----GYEKKTKAAE 310  
 DB 391 -CNSANLRKPEGNK-----NPMVTNVAPPKPGKHGTVDQENENIQPREKQVSQKE 440  
 QY 311 VRNLLATRNPS-----SQDSYIMAKVYVVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPN 366  
 DB 441 -RTVPTTRDPSPGWRNSQDYGKSNYKL-----PQ-----PEDNNMLVPN 479  
 QY 367 LG-----RYITEQPMAFQIIVLRQEFVDSVRDDPYGLPWKEA---VAQ 407  
 DB 480 FNSIDQRENSYPRGESKRAPNSDGTQTQIIPK-----GIVLEPRRIPESETNQPELK 534  
 QY 408 HIAKNP--TDALIPFRPEP-----QVTPTEEHPMHTQIHRDAFSYGAVGPEV 456  
 DB 535 HSAYQPVYTEGIPSPAKEHFPAGRNWTNQOEISPPPEKED-PCRQEBHLPHLSHG-----587  
 QY 457 DSRVIVDLRFWFGATDPEANLLVFQNDVODGYSMP-----QPTFRYPSTASNVARKMMA 512  
 DB 588 -SRVHYYPDPYDPNPDPRENSPYLSRNTWYERDDSPNTMGQPNPHYPMTDPKET-----642  
 QY 513 DMCEVASNLGGVLPSPQPMFMDPLGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAG 572  
 DB 643 -----IPNEEDPIDTGDHEFPQOSR-----WDMEELSFKE 674  
 QY 573 NCTIRTFGENPT 585  
 DB 675 DPTVRHYEGEQYT 687  
 -----  
 RESULT 10  
 ID MTLD\_KLEPN STANDARD; PRT; 382 AA.  
 AC Q9XB6;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17).  
 GN MTLD.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Klebsiella.  
 OC NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KAY 2026;  
 RA Otte S., Lengler J.W.;  
 RT "The mtI genes and the mannitol-1-phosphate-dehydrogenase from  
 RT Klebsiella pneumoniae.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -!- CATALYTIC ACTIVITY: D-mannitol 1-phosphate + NAD(+) = D-fructose  
 CC 6-phosphate + NADH.  
 CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF166095; AAD45386.1; -  
 DR InterPro; IPR000669; Mannitol\_dh.  
 DR Pfam; PF01232; Mannitol\_dh; 1.  
 DR PRINTS; PR00084; MTLDDHGNASE.  
 DR PROSITE; PS00974; MANNITOL\_DHGENASE; 1.  
 KW Oxidoreductase; NAD.  
 FT NP\_BIND 3 14  
 SQ SEQUENCE 382 AA; 41130 MW; D9927B6C068C5344 CRC64;  
 -----

Query Match 3.0%; Score 98.5; DB 1; Length 382;  
 Best Local Similarity 22.0%; Pred. No. 4.6;







Db 108 PPGQALPSAPGVVLG--PEDLPVEVLQFHTSDGILVSAAGTTVVKVWDAKQ---PLT 162  
Qy 184 SLPGIGRPKLSNDPAEDDKENNELYSSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQ 243  
Db 163 EL-----AAHGLVQSAVMSRDGALVGT-----ACKDKQ 191  
Qy 244 -RIRPLPLACHRLK-NAPEYVEVH-----SAENLFHSIYNDKOKKL----- 284  
Db 192 LRIEDP-----RTKPRASQSTOAHENSRRSLAWMGTFWEHLVSTGFNQMREREYKLMWT 245  
Qy 285 -----FTLLNHRT-----RLALTGYEKIKGAAEYVNRLLATRNFSQ--L 324  
Db 246 RFSSALASLTLDTSGLCVPLLPDPSGLLVLAGGERQLCYEVVPOQALSPVTCQVL 305  
Qy 325 DSYIMAKVYVLAGAIGNPQILYNSGSLGQVTRNDLSLPLNGLRYITEQPMACQIVLR 384  
Db 306 ES-----VLRGAAVPRQALVAMGCEVLRVQLSDTAIVPIGYHVPRKAVEFHEDL-- 356  
Qy 385 QEFVDSV-----RDDPYGLPWW--KEAVAQHIKAP-----TDALPIPRDPPEQVTT 430  
Db 357 --FPDTAGCVPATPHG--WAGDNQVQVQVSLNAPACRPHPSFTSCL-VPPAELPDTAQ 411  
Qy 431 PFTEHPWHTQIHRDAFSYGAVGEVDSRVIVDLRWFGATDPEANLLVFNQDVQGYSM 490  
Db 412 PAVMETP-----VGDADASEGFS 430  
Qy 491 PPTFRYRPSPASNV 505  
Db 431 P-PSLTSPTSPSSL 444

## RESULT 14

CCAA\_BACTO STANDARD; PRT: 1257 AA.  
ID CCAA\_BACTO Q45754;  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pesticidal crystal protein cryIIaA (Insecticidal delta-endotoxin  
DE CryIIA(a)) (Crystalline entomocidal protoxin) (142 kDa crystal  
DE protein).  
GN Cry12AA OR CRYXIIA(A) OR CRYVB.  
OS Bacillus thuringiensis  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1428;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.  
RC STRAIN=NRRL B-18244 / PS33F2;  
RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,  
RA Sick A.J.;  
RT "Novel Bacillus thuringiensis microbes active against nematodes, and  
RT genes encoding novel nematode-active toxins cloned from Bacillus  
RT thuringi.";  
RL Patent number EP0462721, 27-DEC-1991.  
CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.  
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
CC OF THE SPORE COAT.  
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
CC N-TERMINUS.  
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L07027; AAA22355.1; -  
DR InterPro: IPR001178; Endotoxin.

DR Pfam: PF00555; endotoxin; 2.  
RW Toxin; Sporulation.  
SQ SEQUENCE 1257 AA; 142265 MW; 3D9888FEC6C0E3981 CRC64;

Query Match 3.0%; Score 98; DB 1; Length 1257;  
Best Local Similarity 20.3%; Pred. No. 30;  
Matches 132; Conservative 88; Mismatches 243; Indels 188; Gaps 36;

Qy 15 SMQINGQIPKNAIHETHTYGNQGVDFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSYA 74  
Db 300 SSPIYQVPYKNNQNTSSSIVPSDLFHYQGLVKLEFSTRTDNDGLAKFT--GIRNTYK 357  
Qy 75 VNAEEGTAVPVVGYHKKNEIEFQKIDRFNVNKGALQVSVVPRNVNQPVT-LDPGAMS 133  
Db 358 -----SPNTHETHYHDFSYNTQSSGNISRGSSNPIDLNPNPIISTCIRNSEYK 406  
Qy 134 APPGSSAISNCKNPHOREFENLSAEAVTRGVCGNTHWTCSTPRIHPWESLPGIGRKL 193  
Db 407 AIAGSSVLVNFKDTQ-----GYAFAQAPTGGA-----WD-----HSFIES----- 442  
Qy 194 SNDAEADDKENNELYSSEARLIGTSTKEFDESIRHTLV-LRSLQDAYKDRQIRPLPLA 252  
Db 443 --DCAPEGHKLNITYTSP----GDTLRDFINV--YTLISTPTINELSTEKIKGP---PAE 491  
Qy 253 CHRLKN-----APEYVHWSAENLFHSIYNDKOKKFLTLTNHRCRTRALTTGGYK 304  
Db 492 KGYIKNOGIMKYKPEYINGAQPVL-----ENOQT LIFE-----HASKTAQYTI 538  
Qy 305 KIGAAEVRNL-----LATRNPSSQLDSYIMAKVYVVLASGAIGNPQILYNSGFSGL 354  
Db 539 RIRYASTQGTGKGYFRLDNQELQTLNIPTSNGYV-----TGNIGENYDLYTIG--SY 588  
Qy 355 QVTPRNDSLIPNLGRYITEQPMACQI-----VL-RQEFV--DSVRDQDPYGLPWKE 403  
Db 589 TITEGNHTL-----QIOHNDKNGWLDRIEFVPKDSLDSP----- 624  
Qy 404 AVAQHIAKNPTDALPIPRDPPEQV---TTPFTEHP--W-----HTQIHRDAFSYGAV 452  
Db 625 -----QDSPPEVHESIIFDKSSPTIWSNKHSHYSHILEG-SYTSQ 665  
Qy 453 GPEVDSRVIVDLRWFGATDPEANLLVFON-DVQDGYSMQPTFRYRSTASNVPRKMM 511  
Db 666 G-SYPHNLLNL--FHPTDPNRNHTIHNNGMDMNDYG-----KDSVADGLNFNKIT 714  
Qy 512 ADMCEVASNLG-----GYLPTSPQFMDPGIALHLAG-TTRIGFDKAT-----TVA 556  
Db 715 ATIPSDAWYSGTITSMHLFNDNNFKTITPKF---ELSELENIITVOVNAFASSAQDTLA 771  
Qy 557 DNNSLVMDFANLYVAGNGTIRTFG-ENPTLTSMCHAIKARSIIINTLKG 606  
Db 772 SNVSDYV-IEQVMKVDALSDEVFGKELKRLKLVNQAKRLSKIRNLLIG 821

## RESULT 15

Y492\_MYCTU STANDARD; PRT: 684 AA.  
ID Y492\_MYCTU Q11157;  
AC Q11157;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 57.3 kDa protein GMC-type oxidoreductase RV0492c.  
GN RV0492C OR MT0511/MT0512 OR MTCY20G.18C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
RT Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.\*;  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.\*;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- COFACTOR: FAD (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.  
CC -!- CAUTION: REF.2 SEQUENCES DIFFER FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 196.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 277162; CAB00953.1; .  
DR EMBL; AE006952; AAK44734.1; ALT\_FRAME.  
DR EMBL; AE006952; AAK44735.1; ALT\_FRAME.  
DR TIGR; MT0511; .  
DR TIGR; MT0512; .  
DR TuberculList; RV0492c; .  
DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
DR InterPro; IPR000172; GMC\_oxred.  
DR InterPro; IPR00205; NAD\_binding.  
DR Pfam; PF00732; GMC\_oxred; 1.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.  
DR PROSITE; PS00624; GMC\_OXRED\_2; 1.  
KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD;  
KW Complete proteome.  
SQ SEQUENCE 684 AA; 72253 MW; 0242535D29E9FB7B2 CRC64;

Query Match 2.9%; Score 96.5; DB 1; Length 684;  
Best Local Similarity 19.2%; Pred. No. 16;  
Matches 123; Conservative 61; Mismatches 224; Indels 233; Gaps 27;  
QY 37 DVFIAAGSPIGATYAKLVEAGLRVVMVEIGNADSEFVAVNAEEGTAVPVPGYHKHNEIE 96  
DB 203 DAVVVGSGAGGAMVARTLARAGLDVVLVEGRR---WTV-----EEFR 242  
QY 97 FQKIDRFVNVTKGALQOVSPVRNQNVPTLDPGAWSPGSSAISNGKNPHQREFENLS 156  
DB 243 STHPVDRYAGLRGAGATVA-----LGRPAVLPMG----- 273  
QY 157 ABAVTRGVGGMs--THWTCSTPRIHPHMEsLPGIKPLSLNDPAEDKWNELYSAEARL 214  
DB 274 -----RAVGGTTVVNSGTCFRPSL-----AVQRRWRDEFLGLADPDQLGRRL--DDAEGT 322  
QY 215 IGTSTKEFDESIRHVLTLVLSQDAYKDRQIRPLPLACHRLKNAP----- 260  
DB 323 LRVAVPVLEIMGRNG---RLLIQAASLKGWRAAIP-----RNAPGRCGCCQCAIGCPS 373  
QY 261 -----EYVENHSAENLPHSIYNDKQKLFLLTNHRCRLALTGG 301  
DB 374 NAKFGVHLNALPOACAAGARIISWARVERILHR----- 406

QY 302 YEKIGAAEVRNLATRNPSQLDSYIMAKVVVVLASGAIGNPQILYNSGFSGLQVTPRND 361  
DB 407 -----AGRAYGVRRRRPDGTTLD--VLADAVVVAAGATETPCLLRRSGLGG----- 450  
QY 362 SLIPNLGRYITEOPMAFCQIVLRQEFVDSVRDDPYGLPWWKEAVAQHIKAK---NPTDALP 418  
DB 451 --HPRLGHNLALHPAT---MLAGLFDDDV-----FAW--RGVLQSAAVHEFHESDGV 496  
QY 419 IPFRDPEPOVTTPTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRW-----FGA 469  
DB 497 IEATSTPPGM-----GSMVFPGYGAEL-----LRWLDRAPIATFGA 533  
QY 470 TDPEANNLLVFNQDYGDSMPQPT--FRY--RPSTASNVRR-----KMMADMCEVAS 519  
DB 534 -----MVADRGVGVTVRSRGETVVRVDIAPGETAKLRVALQAIGRLLFAAGAVEVLT 585  
QY 520 NGGYLPTSPQPMDPGL-----ALHLA-----CTTRIGFDKATTVADNNSLWVD 564  
DB 586 G-----IPGAPPMRSLPELQDVLRRANPRSLHLAAFHPTGTAAAGADEQLCPVDATGLRLG 641  
QY 565 FANLYVAGNGTIRTGFGENPTLTSMCHATKSARSIIINTLKG 605  
DB 642 VEGVWVADASILPSCPVEVNPQLSIMAMALAVADQTVAKVVG 682

Search completed: September 3, 2002, 16:01:24  
Job time: 333 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:49:55 ; Search time 54.14 Seconds  
(without alignments)  
1096.845 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSTEOMLRDYPRSMQING.....IINTLKGGTGDKNGEHRNL 618

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	6.1	573	2 E81385	probable oxidoredu
2	191.5	5.8	615	2 B38575	gluconate 2-dehydr
3	190.5	5.8	579	2 H87451	oxidoreductase, GM
4	187.5	5.7	619	2 AB2003	dehydrogenase chai
5	182	5.5	591	2 H83362	gluconate dehydrog
6	179.5	5.5	529	2 A84260	hypothetical prote
7	176	5.4	722	2 B75608	GMC oxidoreductase
8	171.5	5.2	562	2 JC7628	glucoside 3-dehydr
9	158.5	4.8	748	2 T10651	hypothetical prote
10	148.5	4.5	561	2 E8192	probable oxidoredu
11	148.5	4.5	561	2 AE3094	oxidoreductase Atu
12	138.5	4.2	527	2 E98202	dehydrogenase chai
13	138.5	4.2	527	2 AD3084	dehydrogenase Atu4
14	138.5	4.2	768	2 JC6564	cellobiose oxidase
15	138	4.2	494	2 AD3255	l-sorbose dehydrog
16	130.5	4.0	770	2 S60676	cellobiose oxidase
17	127.5	3.9	578	2 F70736	probable chop prot
18	124.5	3.8	578	2 D93400	probable oxidoredu
19	123	3.7	531	2 B83600	probable oxidoredu
20	121	3.7	1498	2 AF1082	B. subtilis yuka p
21	117	3.6	736	2 A86171	hypothetical prote
22	116	3.5	502	2 T18562	hypothetical prote
23	116	3.5	514	2 B70917	probable zwf2 prot
24	116	3.5	599	2 T19711	hypothetical prote
25	114	3.5	1012	2 T00958	hypothetical prote
26	113.5	3.5	678	2 T05821	hypothetical prote
27	111.5	3.4	1042	2 T16169	hypothetical prote
28	110.5	3.4	589	2 T50698	probable mandelonl
29	110.5	3.4	5170	2 T15348	hypothetical prote

30	109.5	3.3	1125	2	T19193	hypothetical prote
31	108	3.3	1498	2	AG1439	B. subtilis yuka p
32	107.5	3.3	557	2	F83181	probable GMC-type
33	107.5	3.3	567	2	AC0143	choline dehydrogen
34	107.5	3.3	964	2	S48404	probable membrane
35	107	3.3	1018	2	T40253	hypothetical prote
36	106.5	3.2	775	2	B72074	hypothetical prote
37	106.5	3.2	775	2	C81594	hypothetical prote
38	106.5	3.2	775	2	D86549	hypothetical prote
39	106.5	3.2	832	2	H84848	phospholipase D [1
40	105.5	3.2	1611	2	T38236	hypothetical prote
41	105	3.2	499	2	T45749	hypothetical prote
42	104	3.2	509	2	D87452	oxidoreductase, GM
43	104	3.2	585	2	S72824	cholesterol oxidas
44	104	3.2	986	2	T33135	hypothetical prote
45	104	3.2	1917	2	C88728	protein C48A7.1 [1

ALIGNMENTS

RESULT	1
E81385	probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 1
C:Species:	Campylobacter jejuni
C:Date:	31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession:	E81385
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil	
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba	
Nature	403, 665-668, 2000
A:Title:	The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number:	A81250; MDID:20150912
A:Accession:	E81385
A:Status:	preliminary
A:Molecule type:	DNA
A:Residues:	1-573 <PAR>
A:Cross-references:	GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74251.1; PID:g696
A:Experimental source:	serotype O2, strain NCTC 11168
C:Genetics:	
A:Gene:	Cj0415

Query Match

Best Local Similarity	20.6%;	Score	200.5;	DB 2;	Length	573;			
Matches	135;	Conservative	74;	Mismatches	270;	Indels	175;	Gaps	26;

QY	36	VDFYIAGSGPIGATYAKLCVCEAGLRVVMVEIG---AADSFYAVNAEEGTAVPYVPGVHKH	92
Db	8	VDVVTYAGWTGTVAAELTKAGLNLSLGRHMQSTENFNIHDEWRYGINY-----	60
QY	93	NEIEFQKIDIDRFVNVIKGALQQVS---VPVRNQNVPTLDPGAWSAAPPSSAISNGKNPHQ	149
Db	61	-----GLMQDCSKDTVTFRH-----DPSGLALP-----	83
QY	150	REFENLSAEAVTRGVGGMSTHTCSTPRIHPPMESLPGIGRP-----KLSNDPAEDD----	201
Db	84	--YRKMGSFLLGNVNGVAGVHNGWTFREFMPYDFEITLSKQRYGNKGLNDYTLQDMGVT	141
QY	202	-KENNELYSEARLICTSYKEF-----DESIRHTLVLSLQDAYKDRQRI	245
Db	142	YKDMEPYIDRFKTCGVSCEPNPLAEKMGAFRSPYPQPELTKMLKRFESAASNLH	201
QY	246	FRPLPLA-----CHRLKNAPEYVEVHSAENLPHSIYNDKQKK	283
Db	202	TYRLPASNSKGGYTNPDGQDLAPCQVCAVCERFG-----CEYGAASPLNTVTPKAMSTG	256
QY	284	LFTLLNHRCTRILATGGYEKKIGAAAEVRNLLATRNPSQLDSYIM-AKYVVLASGAIGN	342
Db	257	KYTIRTVSNVTQIL-----KKDGKVTGVKVFVDTRT-----MKEYIQPADIVILTSMFNN	306
QY	343	PQILYNSGFGSLQVTP-----RNDSLIPNLG--RYITEQPMACQIVLROEFVDSVR	392
Db	307	AKLLMVSNII-GEQYDPKTKGTGLGRNYCYQMMGCTGAFFDEQFNTF-----MGSGALGTTTS	361

QY 393 DDPYG--LPWKE-----AVAQHIANKPTDALPIPPRDPPEQVTTPTTEHPHWHQIH 443  
Db 362 DDFNGDNFDSKEFLHGAMISVQLGTRPIQSAPLPAGAP-----TWGAER- 408  
QY 444 RDAFSVG-----AVGPEVDSRVIVDLRWFEGATDPANLLVFQNDVQDGYSPQPQTFYR 498  
Db 409 KKAALYNFTRATYVGQ-----GASLPHKNYLSLDPTYKDAFGMPLRLTYN 456  
QY 499 PGTASNVRARKMAD-MCEVASNLGGLYPTSPQFMD-----PGLALHLAGTTRIGEDKA 552  
Db 457 -FTDQDRALHKKITDKTAEVARKMGVKSIGKAYILKDYVVPYOSTHNTGTTMGADRE 515  
QY 553 TTVDANNLSLWDFANLYVAGNCTIRTFGECNPTLFSMCHAKSARSIINTLKGK 606  
Db 516 TSVVNTYQLQHWADNLVYVVGAGNFQHNSGYNPTDVTGALAYRCABGILKYHKS 569  
RESULT 2  
B38575  
gluconate 2-dehydrogenase (FAD) (EC 1.1.1.99.3) catalytic chain precursor [validated] - EMBL  
C:Species: Erwinia cyripedii  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 08-Dec-2000  
C:Accession: B38575; T52554  
R:fum, D.Y.; Lee, Y.P.; Pan, J.G.  
J. Bacteriol. 179, 6566-6572, 1997  
A:Title: Cloning and expression of a gene cluster encoding three subunits of membrane-bound  
A:Reference number: A38575; MUID:98012950  
A:Accession: B38575  
A:Molecule type: DNA  
A:Residues: 1-615 <YUM>  
A:Cross-references: GB:U97665; NID:g2584859; PIDN:AAC45885.1; PID:g2584862  
A:Experimental source: strain ATCC 29267  
A:Note: part of this sequence, including the amino end of the mature protein, was determined  
A:Accession: T52554  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-615 <YU>  
A:Cross-references: EMBL:U97665; PIDN:AAC45885.1  
A:Experimental source: ATCC29267  
C:Genetics:  
A:Start codon: GTG  
C:Complex: heterotrimer, consisting out of cytochrome c chain (T52555), catalytic chain  
C:Function:  
A:Description: EC 1.1.99.3 [validated, MUID:98012950]  
C:Keywords: FAD; oxidoreductase  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-615/Product: gluconate 2-dehydrogenase catalytic chain #status experimental <MAT>

Query Match 5.8%; Score 191.5; DB 2; Length 615;  
Best Local Similarity 21.1%; Pred. No. 1.9e-06;  
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDVFAGSGPTGATYAKLCVAGLRVWVEIGA-----ADSFYVNAEEGTRAVVPYGYH 90  
Db 29 VDADVVGFGWAGAIMAKELTEAGLNVALERGPHRDTYPDGAYPOSIDELT-----YN 81  
QY 91 KKNTEFQKIDRFVNVIKGALQVSVPRNVNPTLDPGAWSPGSSAISNGKNPHQR 150  
Db 82 INKKL-FQ-DLSKSTVTHDASQTAVPYR----- 109  
QY 151 EPENLASAVTRGVGMSTHWTCTSPRIHPHMESLPG-----IGRPKLSNDPAEDD--KE 203  
Db 110 ---QLAAFLPGTGTGAGLHSGVHFVRDVPVELNLSRHYEARYGKGFPEGTIQDFGVS 166  
QY 204 WNEL---YSEARLLGTSTKFEDESIRTLVLSQDAY---KDRQRIERPLP-----IAC 253  
Db 167 VNELEPFDFQAEKVFGTSGSAW--TIKGMICKERKGGNFYAPDRSSDF-PLPAQRRTYSA 223  
QY 254 HRLKNAPVEVWH-----SAENLFHSIYND----- 279  
Db 224 QLFAQAESVGHYPDMPSANTSGPYTNTYGAQMGPFCNGCYGSIYACMYSKASPNVNI 283

QY 280 ----KQKLFLLTHNRCTRRLALTGGEYKKGIGA-----AEVRNLLATRNPSO--LDSYIM 329  
Db 284 LPALQEPKEFELRNNAVYLRVNLTDGKKRATCVTVLDQGREVV--QPADVLILSAQF 340  
QY 330 AKVYVLAGATGNP--QILYNSGFGSLQVTPRNDLSLPLN-GRYTEQPM---AFQCIVLR 384  
Db 341 HNVHLLMSLGGQPNPITNEGTVGRNFAYQNIISTLKALFDKNTTNPFGIGAGGAVD 400  
QY 385 QEFVDSVRDDPYGL---PWNKEVAQHIANKPTDALPIPRDPPEQVTTPTTEHPHWH 440  
Db 401 DFNADNFQHPGYFGVGGSPFW---VNOAGTKPVSGLEPFGKTPN-----WGS 444  
QY 441 QIHRDAFSYAGVGEVDSRVIVDLRWFEGATDPEANNLLVFQNDVQDGYSP--QPTTFYR 498  
Db 445 QWK-----AAVDYNNHHISMDAH--CAHQSYRANLYLDLPNKYKNVTCQPLLRMTFDWQ 496  
QY 499 PSTASNVR-ARKMADMCEV--ASN-----LGGYLTSPPPQMDPGL--ALHLAGTTRIGF 549  
Db 497 D---NDIRMAQFMVGKMRKITEAMNPKMIIG--AKGPGTHFTDTTVYQTHMSGGAIMGE 551  
QY 550 DKATTVDANNLSLWDFANLYVAGNCTIRTFGECNPT 585  
Db 552 DPKTSAVNRVLSQWDVDPNVFVPGASAFPOGLGYNPT 587  
RESULT 3  
H87451  
oxidoreductase, GMC family CC1634 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: H87451  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: H87451  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-579 <STO>  
A:Cross-references: GB:AE005673; NID:g13423038; PIDN:AAK23612.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1634

Query Match 5.8%; Score 190.5; DB 2; Length 579;  
Best Local Similarity 20.2%; Pred. No. 2.1e-06;  
Matches 136; Conservative 83; Mismatches 270; Indels 185; Gaps 28;

QY 18 INGOIPKNAIHETYGNDGVDVFIAGSGPIGATYAKLCVAGLRVWVMEIGAADSFYAVNA 77  
Db 4 LNGRARRK---NTY-----DAIVVSGITGGIAAKELTEKGLKVLVL----- 42  
QY 78 BEGTAVPVPGYHKKNEIEFQKIDRFVNVIKGALQVSVPRNVNPT--LDPGAWSAP 135  
Db 43 ERGPMVRHLEDY-----PTAMLDPMQSKYP 67  
QY 136 PGSSA-----ISNGKNPHQRE--FENLSABAVTRG--VG 165  
Db 68 QGKLPEALNAHYKVQRRTYGTMTETQTHFFVRDDEHPYTEENRFDWI-----RGVHVG 121  
QY 166 GMSTHWTCTSPRIHPHMESLPGIGRPKLSND---PAEDDKWENLYSARLLGTSTKEF 222  
Db 122 GRSLTWGRQSYR-HSPID-FEANAREGIAVDWPIRYEDLAPW---YEHRFVIGVSGQA- 175  
QY 223 DESIRH-----TLVLSQDAYKDRQRI-FRPLPLACHRLKNAPYVSWHSA----- 268  
Db 176 -EGLPHFPDGHVQPPMELNCVEKAFKARSEARFPERRVYICRTAHLTDPTTEQLALGRTK 234  
QY 269 ---ENL-----FHSIYNDKQKKLFTLLTHNRCTRRLALTG-----YKKKIGAAEVR 312

Db	235	CQYNLCIRGCPGAGYYSSNG-----GLIAAERTGNLVIRPNISIVTELIYDERAGRASGV	290
Qy	313	NLLATRNPSQLDSYIMAKVYVYLASGAINGPQILYNSGSLQVTPRNDSLIPN-----	366
Db	291	RIL----DAETRKDBEEFHADVIFLCASALNSAWIMWNS-----TSSRPNGFGNAS	337
Qy	367	--LGRYTEQPMACQIVLRQEFVDSV----RDDPYGLPMWKEAVAOHAKTNPTDALPTP	420
Db	338	DQLGRNVMHHLGAGATGQAPFADMYFSGRRPNGIYVPRFRN-LGDAASKRSDYLRGFG	396
Qy	421	FRDPEQVTTPTTEEHPPWHTQIHRDAFSYGA-----GPEVDSRVIVDLRWFGATDP	472
Db	397	YQGGAGRAT-----WERDGGGGRGFGCAARKAALSQGPWMT-----MGLSGFCGMLP	443
Qy	473	EANNLLVFQNDVQGYSMPOPTFRYRPSTASNVRAKRMADMACEVASNLGGYILPTSPPOF	532
Db	444	YADNRVTLNRDVEDKFGFLPTLTLMNVTMRDNEMARRMDQAAAAEMLEAAGFQNVRAHDNG	503
Qy	533	MDPGLALRLAGTTRIGFDKATTVADNNLSLVMDFANLYVAGNGTIRTFGFGENPTLTSMCHA	592
Db	504	FAPGLGIHEMGTRMGDRPKTSVLNAHNOVHECKNYYVTVDGAAMASASCNVSLTYMALT	563
Qy	593	IKSARSIINTLKG 606	
Db	564	ARAAADHAVRAKRG 577	
RESULT 4			
AB2003 dehydrogenase chain [imported] - Anabaena sp. (strain PCC 7120)			
C:Species: Anabaena sp.			
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120			
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2001			
C:Accession: AB2003			
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, N.			
DNA Res. 8, 2005-213, 2001			
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. PCC 7120			
A:Reference number: AB1807; MUID:21595285; PMID:11759840			
A:Accession: AB2003			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-619 <KUR>			
A:Cross-references: GB:BA000019; PIDN:BA077942.1; PID:g17135396; GSPDB:GN000001			
A:Experimental source: strain PCC 7120			
C:Genetics:			
A:Gene: alr1576			

Query Match	5.7%	Score 187.5;	DB 2;	Length 619;
Best Local Similarity	19.7%;	pred. No. 3.9e-06;		
Matches 139;	Conservative	80;	Mismatches 242;	Indels 243; Gaps 29;
Qy	29	ETYGNDGVDFIAGSPGIGATYAKLCVBAGRURVVMVEIGAADSFYAVNAEBGTAVPYVPG	88	
Db	12	DTVSKTVYDAVIVGTGVAGAIKAVELSSQGRKILIIIEATV-----	51	
Qy	89	YHKKNEIE-FQKDDIRFVNVITKGALQGVSPVRNQNVPTLDPGANSAPPGSSA-----	140	
Db	52	-HKDLTLAGFOSYVDTFYKAVD-----KNPNS-----YPANSVNOSPSTDYN	92	
Qy	141	---ISNGKNPHOREFENLSAEAVTRGVCGMTHWCSTPRHP---PMESLPGTIGRPKLS	194	
Db	93	DYFTEQGPMP-----LAGSYTRVLGGTTHHWEAKTPRMLPDFKLUSSYVGGQL-----	140	
Qy	195	NDPAEDDKENWELLYSEAERLIGTSTK-----EPDESIRHTLVLRSLQADYKDRORI	245	
Db	141	-DWPIDYHDLEPPYRYKAEHMGVCGDVDEQRAIGLEFPQD--YVFPMEKLPPLYLDQKVI	197	
Qy	246	-----FRPLPLACHRLKNAPEVEMHSAENLF-----HSIYND	279	
Db	198	EKVGNTNVELYGKHTLSFSFTFPOARNGVPN--PKY-----DQGNLFPVDPGVTSVHPVOQGE	252	

Qy	260	K-----OKKFLTLLTNHRCTRLALTGGEYKKIGAAAEVRNLLA	316
Db	253	: :	312
Qy	317	TNNSSQLDSIMAK--VYYLASAIGNPQILYNSGSGLOVTPRNDLSLIPNGRYITEQ	374
Db	313	: :	360
Qy	375	PMAFCQIVLRQEFVSDVRDDPYGLPWPKKEAVAQHIKNPDALPIPPDPPE-----PQ	427
Db	361	-----MDHPFTLAW-----ALM-----PEVTGMRGPL	363
Qy	428	VTTPF-----TEEPWHQTQIHRDAFSYGAVGPEDVSRIVD-----	463
Db	384	: :	443
Qy	464	-----LRWTGATDPEANNLLVFONDVGYSMPQTFPRYRPSTASNV--RAKMMADM	514
Db	444	RISRQLLAFMCLELLPEYGNRVITIDPRHKDLG-----NYRPVINFNLDYSRRTLAYT	497
Qy	515	CEVA:--SNLGG-----YLPTSPQPMDPCLAL-----HLAGTTIRIGFDKATTVADNN	559
Db	498	RKVRSVIFERGAEDYHYDQDPAYEFEFEGGVYKGGNHFGTHMGTTPLNSVVDSY	557
Qy	560	SLVWDFAFLYVAGNGTIRTGFENPTLTSMCHAIKSARSIINTL	603
Db	558	LRSWDHKNLFVLVGAGSMPTIGSSNTTLIAALSFTAEHMLQEL	601
RESULT	5		
H83362		glucanate dehydrogenase PA2265 [imported] - Pseudomonas aeruginosa (strain PAO1)	
C:Species:		Pseudomonas aeruginosa	
C:Date:		15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000	
C:Accession:		H83362	
A:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.			
: Lory, S.; Olson, M.V.			
Nature		406, 959-964, 2000	
A:Title:		Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa	
A:Reference number:		A82950; PMID:2043737	
A:Accession:		H83362	
A>Status:		preliminary	
A:Molecule type:		DNA	
A:Residues:		1-591 <STO>	
A:Cross-references:		GB:AEO04652; GB:AEO04091; MID:99948287; PIDN:AAG05653.1; GSPDB:GN	
A:Experimental source:		strain PAO1	
C:Genetics:			
A:Gene:		PA2265	

Query Match	5.5%	Score 182;	DB 2;	Length 591;
Best Local Similarity	20.6%;	Pred. No. 9.7e-06;		
Matches 138;	Conservative 66;	Mismatches 261;	Indels 204;	Gaps 23;
Qy	36	VDVFAGSGPGATYKACLVCEAGLRVAVVEICAAADSFYAVNAEEGTATVPVPGYKHKNEI	95	
Db	8	VDVVVFGWGTGAILAKELTEAGLNVVALERGENRDTYPDGAYPNTL	57	
Qy	96	EFQKDIDRFVNVKIGALQQVSPVRNQNVPTLDPCAWSAPPQSSAISNGKNPHQREFENL	155	
Db	58	TYNTRGKLFQNLKSTV	90	
Qy	156	SAEAVTRGVGGMSTHWTCSTPRIHPMPSLPG	205	
Db	91	SAFLPGDGVGGAGLHWSGVHFRIMPEELRLSHYBERRYCKKFIPIGWTIQDYGVSYEELE	150	
Qy	206	ELYSAEARLIGTSTKEFDESIRHTL	246	
Db	151	PHFDFAEKVFGTS	202	
Qy	247	R	276	

Db 203 RKAABELGLHPYDLPAANASGPTWNPYGVQMGPCNFCEGFCGYACMYTKASPNL--NIL 260  
QY 277 NDDKQKFLTLTNHRCRLALITGGYEKKIGA-----AEVRNLLATRNPSQLDSYIMARV 332  
Db 261 PALRQTPFLFELRANCNVLKVNLDSDGRQATGYVYVDAQGREIVQP-----AKL 308  
QY 333 YVLASGATGNPOILYNSGSLQVTPRNDLSLIPNIGRITTEOPMAFCQIVLRQEVDSVR 392  
Db 309 VIISAFQPHNVHRLLLSGT-GKPYDPTGEGV--VGKNFAYONMA-----TIKAFPDKDVH 361  
QY 393 DDPY-----CLPWKKEAVAQHIAKNPTDLPPIFPROPE 425  
Db 362 TNPFVGTGGGVAVDDFNADNHDHGLFVGSGPMW-----VNQAGSKPIGLAVPPGTPS 417  
QY 426 PQTTPPTEEHFWHT---QIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPEANNLLVFQN 482  
Db 418 -----WGSQWQAVKDAYTH-----TVSMDAHGSNWTYRDNYLDLDP 454  
QY 483 DVODGYSMP--OPTPRY-----RPTASNVRARKMMADMCVEVA-----SNLGGYLPSTPP 530  
Db 455 TYKDAYGQPLLRMTFDWKDNEIRMSRYVTEHMRKTAEAAMPKATSVSVKNEFGDHENTRVY 514  
QY 531 QPMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGFGENPTLTSNC 590  
Db 515 Q-----TTHLLGGAIGSDPKTSVLNRYLOSVDVHNVFVVGASAFPOGTGYNPTGLVAA 568  
QY 591 HAIKSARSI 599  
Db 569 LAYWSAKAI 577  
  
RESULT 6  
A84260  
hypothetical protein Vng1035c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: A84260  
R:Ng, W.V.; Kennedy, S.P.; Maharas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
Leithaus, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84260  
A:Accession: A84260  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-529 <STO>  
A:Cross-references: GB:AE004437; NID:g10580587; PIDN:AAG19445.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG1035C  
  
Query Match 5.5%; Score 179.5; DB 2; Length 529;  
Best Local Similarity 22.1%; Pred. No. 1.3e-05;  
Matches 143; Conservative 53; Mismatches 236; Indels 215; Gaps 28;  
  
QY 37 DVFTAGSGPIGATYAKLCVEAGLRVWVEIGAADSFYAVNAEEGTAVPYVPGYHKKEIE 96  
Db 10 DVCVIGAGPAGGLIADRLASDGHVSVVLEAG--PRFDAADR----- 50  
  
QY 97 FOKDIDRVNVIKALQOVSVVVRNQNVPTLDPGWS--APPGSSAISNGKNPHOREFN 155  
Db 51 -----RMERSIRPAHGPAVY-----WEMGGPRDAYSTGD-----RYVPL 85  
  
QY 156 SAEAVTRGCGMSTHTWCTSTPRIHP-----MESLPGICGRPKLSNDPAEDDKENNELYSAE 212  
Db 86 NAARY-KVGSGSTLHWQGNWRLHQEDPRLASATGVGA-----DWPPDYDFLKPYAAAE 139  
  
QY 213 RL1---GTSTKEFDESIRHTLVLSLODAYKDRQRIFRPLPLACHRLKNAPEYVEWHSAAE 269  
Db 140 SALGVSGASDNPFPAPPREOPHPQAFPPSYSD--SLFAD-----ACESLGIAI----- 185

QY 270 NLFHSIYNDKQKKLFTLLTNHRCRLALTG-----GYE-----K 304  
Db 186 ---HSVPN-----AKLSAGRETRRACVGYGTQCPVPCPGAKYDATVHVDRATDAGAR 234  
QY 305 KIGAAEVRNL-----LATRNPSSQLDSYIMAKYVYVVLASGAIGNPOIL----- 346  
Db 235 VIDEAPVORLEHDAAGDRVTGAVYATPDGTTTHRS---ATEFVLAAGGIETPRLLLSDS 291  
QY 347 --YNSGF---SGLQTPRNDLSLIPNLGRIITEOPMAFCQIVLRQEVDSVRDPYGLPMW 401  
Db 292 DRYPDGLANSGL-----VGRY-----FMD----- 311  
QY 402 KEAVAQHIAKNPTDALPIPRDPEQVTPPTFEHPHWHTQIHRDAFSYGAV-----G 453  
Db 312 -----HLFAGAGGTLDPETRQNHVGFNT--TESHQY---DRPDGSRGAIKLEFLNYAG 360  
QY 454 PEVDSRVIVDLRFWGA-----TPEANNLLVFONDVODGVSMPOQ 493  
Db 361 FSPAEMALSGDDWDGAMCDRIDASGTHIAGVGLVEQQQPRPENRVRHPERTDVHGNVP 420  
QY 494 TFRYRPTASNVRARKMMADMCVEASNLGGYLP-TSPQFMDPGLALHLAGTTRIGFDKA 552  
Db 421 DVVMSLSAYERTIERANEIQREILTELCADIEWTVGPE--DTGPAFHHMGTTRMGTDPA 478  
QY 553 TTVADNNSLVWDFANLYVAGNGTIRTGFGENPTLTSMCHAIKSARSI 599  
Db 479 ESVVDPRLRTHDLSNLSSVSSVFPPTACAMNPTLTIAALAKAADI 525  
  
RESULT 7  
B75608  
GMC oxidoreductase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: B75608  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: B75608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-722 <WHI>  
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12230.1; PID:g646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0127  
A:Map position: 2  
  
Query Match 5.4%; Score 176; DB 2; Length 722;  
Best Local Similarity 20.9%; Pred. No. 3.8e-05;  
Matches 139; Conservative 77; Mismatches 239; Indels 210; Gaps 33;  
  
QY 14 RSMQINGQIPKNAIHTETGNDG---VDVFTAGSGPIGATYAKLCVEAGLRVWVEIGA- 68  
Db 186 RNSAITPVVPO-----DGEVLEADAVVVGSGGGVIAARLAQAGKRVVLEAGGH 236  
  
QY 69 --ASDFYAVNAEEGTAVPYVPGYHKKEIEFQKIDIDRFVNVYIKGALQOVSVVVRNQNVPT 126  
Db 237 YHEAHFOGRELLAAYQTLVYRGYH-----PTADGNV-T 268  
  
QY 127 LDPGAWSAPOGSSAISNGKNPHOREFNLSAEAVTRGVGGMSTHTWCTSTPRIHPMESLP 186  
Db 269 LVAGA-----NLG-----GGSTVWNSV-----PPRDI- 293  
  
QY 187 GIGRPKLSN-----DPAEDDKENNELYSEABRLIGTSTKEFDESIRHTLVLSLQDA 238  
Db 294 ---RQRWASEHGLSDVADPGVD---RHIDAVLERM-GVSEQCSDHNGPHQ---RLVEGA 342  
  
QY 239 YKDRQIRFRPLPLACHRLKNAPEYVEWHSANLFHSIYNDD---KQKKLFT----- 286

```

144 YEDLAPW---YDYVERPAGIAGTQE-----GLDILPDG-----EFLPPIPLMC-----183
QY
257 KNAPEYVEWESAENLFHSIYNDKDKQKLLFTLLNHRCTRLLATGGYEKKIAGAEVRLILA 316
Db
184 -----VEEDAARKI-----KEAF-----GGORHLIHSRVANITQ 212
QY
317 TRNPSSOLDSYIMAKYVVLASGAIGNPOILLYNSGFSLOVTPRNDSLIPNLGRYITEQPM 376
Db
213 PKPEQNVNCQYRNKWC-----LGCP-----YGAYFTQSATLPAAVATGNL-----TLRPF 259
QY
377 AFCCOVL---RQ-----EFVDS-----VRDDPYGLPW-----WKEA 404
Db
260 SIVSOVLVLDQRQARGVEVIDAETHVEHVEYADVIFLNASTENTWILMNSATDWEVG 319
QY
405 VAQIIAKNPDTALPIPR-----DPEQVPTTTEHPWHQTQHR-----DAP 447
Db
320 LGSSSGELGHNVMDHHFRCGASGEVEGLDKYYFGRRPAGFYIPRPNVNGDEQORSYVRGF 379
QY
448 SY-GAVGPEVDSRVIVDLR-----W-----FGATDPEANLLVFOVD 486
Db
380 GYQGAASREGWDEIAELNIGADLKQALTOPGGWTIGMTGFGEMLPDHNRLSLDHSVRD 439
QY
487 GYSMPOTFFRYRPTASNVBRKMWADMC-----VASNLGYLPTSPPFQMDPGL 537
Db
440 KWGLPVLISIDVELKQNERDMRMDVQDAVDLLEAGVKNVKGVGYA-----PGM 490
QY
538 ALHLGAGTRIGFDKATVADNNSLVDFDANLYVAGNGTIRTGEGENPTLTSMCHAIKSAR 597
Db
491 GIHEMGFTARMGRDKPTSVLNSHQNVQDAPNVFTDGACTMTSSSCVNPSTLYMAL/TARAVD 550
QY
598 SIINTLKG 606
Db
551 YAVEELKRG 559

```

Db 551 YAVEELKRG 559

RESULT 9

Tl0651

hypothetical protein T5F17.20 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Sep-2000

C/Accession: Tl0651

R/Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell

submitted to the Protein Sequence Database, June 1999

A/Reference number: Z16533

A/Accession: Tl0651

A/Molecule type: DNA

A/Residues: 1-748 <BEV>

A/Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20

A/Experimental source: cultivar Columbia; BAC clone T5F17

C/Genetics:

A/Gene: ATSP:T5F17.20

A/Map position: 4

A/Introns: 70/3

C/Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160

```

Query Match          4.8%; Score 158.5; DB 2; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00089;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps

QY 37 DVFAGSGPIGATYAKLCVEAGLRVVMVEIGAAISFVAVNAEAGTAVPVVPGVHKRNEIT 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 DAVVVGSGGGVAAANLAKAGLKVLVLEKG---NYFTAHDYSGLEVPSKMLELYEKG--- 292

QY 97 FQKIDIRFVNVIKCALQOVSVPVYRNQNVPTLPDGAWSAPPGCSAISNGKNPHQRFENILS 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 -----GLLTV-----DGK-----FMLLA 306

QY 157 AEAVTRVGVMGSMTHWTCSTPRIHPHMESLPGIGRPKLSNDPAEDDKENWELYSAPRLTIG 216
   || || || || || || || || || || || || || || || || || || || || ||
Db 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----OEW-----SEGSKINF 340

QY 217 TSTKEF---DE-SIRHTLVRLSLQADYKDRORIFRPLPLACHRL-----KNAPE--- 261

```

Db 341 FSGQYQSAMDEVTIRIGVTERCVKHGFQ-----OVLKRGGERLGLQVESVPRNSPEDH 395  
QY 262 -----YVEWHSAENLFHSIYNDKQKLLFLLTHNRCTRLL-----TGGYEKK-----I 306  
Db 396 YCGLGCGYGRACAKNGTDOTWLVDAVENCAGVILTGKAERFVLVDNTSSNERKKRCVGV 455  
QY 307 GAAEVNLLATRNPSQLDSYIMAKYVYLASGAIGNPQILYNSGSLQVTPRNDSLIPN 366  
Db 456 FASSVGGKIGKK-----FIIEARVTVSSAGSLTTPPLMLSSGLKN-----FN 497  
QY 367 LGR-----YITQPMAFQOIVLROEFVDSR--DDPYGLPWWKEAVAQHIKN 412  
Db 498 IGRNKLKHPVLTWGTGYFPKSEFSGKMEYEGGIITSVHMD-----TESGCKRAILEN 550  
QY 413 PTDALPIPRDPPOVPTTTEHPHWTQIHRDAFSYGAAGVPEVDSRVILVDRWFGATDP 472  
Db 551 PLIG-PASVAGLSPWVS-----GPDLMKRMK-----YGT-- 580  
QY 473 EANNLVFONDYQDYSM--POPTFRYRSTASNVRAKMMADMCEVAS---NLGGY--- 524  
Db 581 --AHLFALVRLDLSGSEVMENEVYRTTKKDRNLRAGLRQALRVSAAGAVEVGTYRSD 638  
QY 525 -----LPTSPQFMDPGLAL-----HLAGTRIGFDKATTVADNN 559  
Db 639 GQMKCEAITKAMEEFLDEVDAGVGVTGKGYWTYFSAHQMGSCRMGVTAEEGALDEN 698  
QY 560 SLVWDFANLYVAGNGTIRGTGFENPTLTSCHAIKARSIIINTLKGTT 607  
Db 699 GESWEAEGLFVDCGSLPSAVGNVPMITIQSTAYCISSKIVDSLQNK 746

RESULT 10  
E98192  
probable oxidoreductase chain CJ0415 [imported] - Agrobacterium tumefaciens (strain C58,  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: E98192  
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: E98192  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-561 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_980  
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;  
Best Local Similarity 19.1%; Pred. No. 0.0033;  
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;  
QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEITE 96  
Db 7 DAIVVSGISGGWAAKELTKGLKVLML-----ERGNIEHTYD----- 46  
QY 97 FOKIDRFVNVIKGALQOVSVPRVNONVPTLDPGAWSPGSSAISNGKNPHOREFENIS 156  
Db 47 -----DKEAWDYPHNRNRAQEMKAKYPVLSRDYL 78  
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPPMESLPGIGRPLK 193  
Db 79 LEEATLGMWADQETPYVEEKRFDFWFRGYHVGGRSLLMGRQTYRW--SQTDFEANKDGI 136  
QY 194 SND---PAEDDKWENELSEARLGTGTKEFDESIRHTLVLSLQDAYKDRQIRPRPLP 250  
Db 137 AVDWIRIEDVSPW---YDYVERFAGISGS-----REGLDILPDGE-FLPPIP 180

QY 251 LACHRLKNAPEVVEVHSAENLFHSIYNDKQKLLF---TLLNHRCTRLLALT----- 299  
Db 181 L-----NEVEDQASRL-----KKAFTGRHLINSRCANITQELPDQERTP 221  
QY 300 -----GGY-----EKKIGAAEVRNLLATR-----N 319  
Db 222 COFRNKRCLGCGFGYFSTQASTLPAAVATGNLTLPFPFSIVKEILYDKDKKARGVEIID 281  
QY 320 PSSQLDSYIMAKYVYLASGAIGNPQILYNSGF---SGLOVTPRNDSLIPNLGRYTE-- 373  
Db 282 AETNLTYETADIIIFLNASTLNSWVLMNSATDVVEGGL-----GSSGELGHVMDHH 335  
QY 374 -----QPMAFQOIVLROEFVDSR---DDP-----YGLP----- 399  
Db 336 FRMGATGOVDFGDEFYFKRRRPAF-----YIPRFRNTGDKRKYLRGFGYQGSASR 387  
QY 400 --WKEAQAHTAKNPTDALPIPRDPPEPOVPTTTEHPHWTQIHRDAFSYGAAGVPEVD 457  
Db 388 SWEREIAELNIGADYKDTL-----TEPGGW--TIGMTAF--GEMLYPHD 428  
QY 458 SRVIVDLRFWFGATDPEANLLVFONDVQDYSMPQPTFRYRSTASNVRAKMMADMCE- 516  
Db 429 NRKLD-----HDKKDKWGL-----PVLSMNVEMKQNELDMRD 462  
QY 517 -----VASNLGGYLTSPQFMDPGLALHLAGTRIGFDKATTVADNNS 560  
Db 463 MYNDAVENFEAVGINKVPSRGTYA-----PGMGIHEMGRTARMGRDPTSVLNGNN 513  
QY 561 LWYDFANLYVAGNGTIRGTGFENPTLTSCHAIKARSIIINTLKG 606  
Db 514 QVMDAPNVFVTDGACMTSASCYNPSSLTYNALTARAAEFAVSRKKG 559

RESULT 11  
AE3094  
oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AE3094  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AE2577; PMID:11743193  
A:Accession: AE3094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-561 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAL45171.1; PID:g17742847; GSPDB:GN00187  
C:Genetics:  
A:Gene: Atu4377  
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;  
Best Local Similarity 19.1%; Pred. No. 0.0033;  
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;  
QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEITE 96  
Db 7 DAIVVSGISGGWAAKELTKGLKVLML-----ERGNIEHTYD----- 46  
QY 97 FOKIDRFVNVIKGALQOVSVPRVNONVPTLDPGAWSPGSSAISNGKNPHOREFENIS 156  
Db 47 -----DKEAWDYPHNRNRAQEMKAKYPVLSRDYL 78  
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPPMESLPGIGRPLK 193  
Db 79 LEEATLGMWADQETPYVEEKRFDFWFRGYHVGGRSLLMGRQTYRW--SQTDFEANKDGI 136





Db 109 VEH-----EDGVSPAWPVSYAEAEFFVYAEAEERLFGV-RGRAGDDPTPEPPRSAPYMH 158  
Qy 209 SEA---ERLIGTSTKDFESIRHTLVLSLQDAYKDRQIRPLPLA-----CHRLKN 258  
Db 159 APIPEVIGRAKGER-----LGLRFH-----MPSAIDYGGGLCRRGT 201  
Qy 259 APEYVHSAENLFHSIYNDKQKLF-----TLTNHRCRTRALTGGYKKIGAA 309  
Db 202 CDAFCVREDAGK-----DAETRLRLPALRHPNVSLTGARVRL-IADGDGKHIVAV 252  
Qy 310 EVRNLLATRNPSQLDSYIMAKVYVILASGATGNPOILYNSGSLQVTPRNDSLIPN--- 366  
Db 253 EIER-----AGEITT-IEAPFLVLSAGAINSALILRSA-----DEKKNGLA 294  
Qy 367 -----LGRVITEQPM-----FCQIVLROEFVDSVRDDP----- 395  
Db 295 NSSGVVGRYLMNHLSGLMLPFTINTTRPKTMSLNFDDFGTGGDGAARGNVQMLGNI 354  
Qy 396 -----YGLPMWKEAVAQHIANKPTDALPIPRDPEQVTTTFTTEHPWHTQIHRDAFSY 449  
Db 355 QGPMIRAAYPWMPRLANLARHSVDFLVM-----SED----- 387  
Qy 450 GAVGPEVDSRVLDLWRTGADPDPEANLLVFNQDVODGYSPQPTFRYRPTASN-----V 505  
Db 388 ---TPKYDSRV---KPMGKNGAE-----LIYRPGDREAHQRFV 419  
Qy 506 RARKMADMCEVASNIGGYLPTSPQPMFDPGLALHLAGTTRIGFDKATTVADNNSLVWDF 565  
Db 420 RHMRSLLRKNFPVVLGHSGFIEAPS-----HQCCTVRMGDDPKAALNALCQTYDH 471  
Qy 566 ANLYVAGNGTIRTGGENPTLTSCHAIK 594  
Db 472 PNLYVVDAGFFPSSAALNPALTVAQAALR 500

## RESULT 14

JC6564  
cellobiose oxidase (EC 1.1.3.25) precursor - white-rot fungus (Trametes versicolor)  
N:Alternate names: cellobiose dehydrogenase  
C:Species: Trametes versicolor (white-rot fungus)  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 19-May-2000  
C:Accession: JC6564  
R:Dumonceaux, T.J.; Bartholomew, K.A.; Charles, T.C.; Moukha, S.M.; Archibald, F.S.  
Gene 210, 211-219, 1998  
A:Title: Cloning and sequencing of a gene encoding cellobiose dehydrogenase from Trametes  
A:Reference number: JC6564; MUID:98248903  
A:Accession: JC6564  
A:Molecule type: mRNA  
A:Residues: 1-768 <DOM>  
A:Cross-references: GB:AF029668  
A:Experimental source: 52J, ATCC 20869  
C:Comment: This is a hemoflavoenzyme that oxidizes cellobiose and reduces iron(III) and  
C:Genetics:  
A:Gene: cdh  
A:Introns: 17/1; 103/2; 140/1; 179/1; 226/3; 281/3; 306/1; 324/2; 534/3; 577/2; 593/3; 6  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-768/Product: cellobiose dehydrogenase #status predicted <MAT>  
F:80,128/Binding site: heme iron (Met, His) (axial ligands) #status predicted

Query Match 4.2%; Score 138.5; DB 2; Length 768;  
Best Local Similarity 19.2%; Pred. No. 0.032;  
Matches 128; Conservative 83; Mismatches 228; Indels 229; Gaps 33;

Qy 37 DVFTAGSGPICATYAKLCVEAGLRVWVVEIG-----AADSFYAVNAEGTAVPY-VPG- 88  
Db 231 DYIVVAGPGLVLTADRLSEAGKVKLLERGGPSTAETGGTYDATWAKSANLTKFDPGL 290  
Qy 89 ----YHKNETEFOKDIDRFVNLTKGALQQVSVPRNVNQTLDPGAWSAPPSSAISNG 144  
Db 291 FETLTTDTPNFWCKDNTFFAGCLLG-----GGTSV-NG 323

Qy 145 K---NPHOREFENLSAEAVTRGVGMSTHTWCSTPPIHPPMESLPGIGRPKLSNDPAEDD 201  
Db 324 ALWYPNSRDFSTAS-----GWPSSWSNHQPTFDKLRQLPS-----TDHPSADG 368  
Qy 202 KENNELYSEA---ERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRF-----RP 248  
Db 369 QRYLEQSATVQQLISGQYSQI-----TINDNPDSDKHVGFSAFDFLNGORA 417  
Qy 249 LPLACHRLKNAPEVHWSAENLFHSIYNDKQKLFLLTNHRCRTRALTGGYKKIGAA 308  
Db 418 GPVA-----TYFETALARKNF---VTKDN---VLVTQVIRNGSTILGVRTN-DNTLGP 463  
Qy 309 AEVRNLLATRNPSQLDSYIMAKVYVILASGATGNPOILYNSGFSG---LQVTPRNDSLIP 365  
Db 464 DGIVPL-----NPNGRV-----ILSGSGFTPRILFQSGIGTDMLOTVQSNQAQAA 510  
Qy 366 NLGRYITEQPMACQIVLRQEFVDSVRDDPYGLPMWKEAVAQHIANKPTDALPIPRDPE 425  
Db 511 NL-----PP-----QSEWID-----LP-----VGQSVSDNPS----- 532  
Qy 426 PQVTPPTTEHP-----WHTQIHRDAFSY-----GAVGPEVDSRVLDLRW 466  
Db 533 ---INLVFT---HPSIDAYDNWADWNSNPRADAQQYLOSRSGLVAGASPKLNF-----WRA 583  
Qy 467 FGATDPEANLLVFQNDVQGYSPQPTFRYRPS-----TASN 504  
Db 584 YGSGDITR---YAQGVTRPGAAASVNTSVAYNASEIFTITLYLSNGIOSGRIGVDAALN 640  
Qy 505 VRA-----RKMADMCEVASNLGGYLPSPQFM-----DPG--LA 538  
Db 641 AKALVNPWLNTSVDKTVLLQALHDVTSTKMNVPGLTMITPDNTMLEQYVAAYDPATMCS 700  
Qy 539 LHLAGTTRIGFDKATTVADNNSLVWDFANLVAGNQTTRTGFGENPTLTSCHAIKSARS 598  
Db 701 NHWVGAARKMGTSSTAVVDENAKVFNTDLFIVDASIIPNIGNPQGVLM-SAAEQAVS 759  
Qy 599 IINTIKGG 606  
Db 760 RILALAGG 767

## RESULT 15

AD3255  
l-sorbose dehydrogenase (FAD) (EC 1.1.99.-) [imported] - Brucella melitensis (strain  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AD3255  
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov  
; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AD3255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-494 <CUR>  
A:Cross-references: GB:AE008917; PIDN:AAL51207.1; PID:gl7981893; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI0025  
A:Map position: 1  
C:Keywords: oxidoreductase

Query Match 4.2%; Score 138; DB 2; Length 494;  
Best Local Similarity 20.4%; Pred. No. 0.017;  
Matches 125; Conservative 67; Mismatches 240; Indels 182; Gaps 23;

Qy 37 DVFTAGSGPICATYAKLCVEAGLRVWVVEIGAAADSFYAVNAEGTAVPYVPYGHKKNEIE 96  
Db 6 DIVIIGSIGGATWAGLAAGSADILILEAGARLADRPENRPAI--FQGFPRFRELW 63



**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:05:10 ; Search time 31.6 Seconds  
(without alignments)  
477.690 Million cell updates/sec

Title: US-09-856-327-2  
Perfect score: 3284  
Sequence: 1 MSLSSTEQLRDPYPRSMQING.....IINTLKGGTDGKNTGHEHNL 618

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2.6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.5	37.5	623	1 US-08-734-925-2	Sequence 2, Appl1
2	1210	36.8	622	4 US-09-305-381-2	Sequence 1, Appl1
3	1091	33.2	566	4 US-09-023-731-1	Sequence 12, Appl1
4	192	5.8	51	4 US-09-023-731-12	Sequence 9, Appl1
5	191.5	5.8	615	4 US-09-297-937C-9	Sequence 13, Appl1
6	191.5	5.8	1276	4 US-09-297-937C-13	Sequence 11, Appl1
7	130.5	4.0	58	4 US-09-023-731-11	Sequence 2, Appl1
8	118.5	3.6	785	3 US-09-265-108-2	Sequence 2, Appl1
9	118.5	3.6	785	4 US-09-479-264-2	Sequence 14, Appl1
10	115	3.5	50	4 US-09-023-731-14	Sequence 15, Appl1
11	108	3.3	35	4 US-09-023-731-15	Sequence 12, Appl1
12	104.5	3.2	1876	2 US-08-609-049A-12	Sequence 6, Appl1
13	104.5	3.2	1876	4 US-09-170-996-12	Sequence 6, Appl1
14	102.5	3.1	61	4 US-09-023-731-6	Sequence 5, Appl1
15	100.5	3.1	806	3 US-08-549-515-5	Sequence 11, Appl1
16	100.5	3.1	806	3 US-08-549-515-11	Sequence 5, Appl1
17	100	3.0	664	1 US-08-485-284A-5	Sequence 2, Appl1
18	98.5	3.0	543	3 US-09-199-229-2	Sequence 2, Appl1
19	98.5	3.0	543	4 US-09-443-087-2	Sequence 2, Appl1
20	98.5	3.0	543	4 US-09-687-298-2	Sequence 2, Appl1
21	98	3.0	1257	1 US-08-049-783-2	Sequence 6, Appl1
22	98	3.0	1257	1 US-08-158-232-6	Sequence 6, Appl1
23	98	3.0	1257	1 US-08-304-626-6	Sequence 6, Appl1
24	98	3.0	1257	1 US-08-316-301A-6	Sequence 6, Appl1
25	98	3.0	1257	2 US-08-611-928-6	Sequence 6, Appl1
26	98	3.0	1257	3 US-09-173-891-6	Sequence 6, Appl1
27	98	3.0	1257	4 US-09-076-137-6	Sequence 6, Appl1

28	98	3.0	1257	5 PCT-US92-03624-6	Sequence 6, Appl1
29	97	3.0	3170	2 US-07-642-734C-5	Sequence 5, Appl1
30	97	3.0	3170	3 US-08-439-009A-5	Sequence 5, Appl1
31	96.5	2.9	1876	2 US-08-609-049A-28	Sequence 28, Appl1
32	96.5	2.9	1876	4 US-09-170-996-28	Sequence 28, Appl1
33	95.5	2.9	382	1 US-08-186-833-2	Sequence 2, Appl1
34	93.5	2.8	903	3 US-08-804-439A-22	Sequence 22, Appl1
35	93.5	2.8	903	3 US-08-720-229-2	Sequence 22, Appl1
36	93.5	2.8	904	4 US-08-632-537-1	Sequence 1, Appl1
37	93.5	2.8	904	5 PCT-US96-05316-1	Sequence 1, Appl1
38	93.5	2.8	904	6 5244792-4	Patent No. 5244792
39	93	2.8	604	2 US-08-746-283-32	Sequence 32, Appl1
40	93	2.8	604	2 US-08-746-257A-30	Sequence 30, Appl1
41	93	2.8	605	1 US-08-333-802-2	Sequence 2, Appl1
42	93	2.8	4545	2 US-08-804-227C-14	Sequence 14, Appl1
43	92.5	2.8	4472	2 US-08-804-227C-2	Sequence 2, Appl1
44	92	2.8	844	1 US-07-731-157A-6	Sequence 6, Appl1
45	92	2.8	844	2 US-08-541-780-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1  
US-08-734-925-2  
; Sequence 2, Application US/08734925  
; Patent No. 5712139  
; GENERAL INFORMATION:  
; APPLICANT: NISHIMURA, IKUKO  
; APPLICANT: OKADA, KIMIHARU  
; APPLICANT: MINAMITHARA, TOMOYUKI  
; APPLICANT: KAWAI, GENSHIRO  
; APPLICANT: KOYAMA, YASUJI  
; APPLICANT: SUZUKI, MASARU  
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE  
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/734,925  
; FILING DATE: 22-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/568,428  
; FILING DATE: 06-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 7126-002-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 623 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-734-925-2



; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MICROSOFT WORD 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,731  
; FILING DATE: 13-FEB-1998  
; PRIOR APPLICATION DATA: JP 29275/1997  
; FILING DATE: 13-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNETH H. SONNENFELD  
; REGISTRATION NUMBER: 33,285  
; REFERENCE/DOCKET NUMBER: 3479-4000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 566  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-023-731-1

Query Match 33.2%; Score 1091; DB 4; Length 566;  
Best Local Similarity 42.6%; Pred. No. 4e-104;  
Matches 248; Conservative 84; Mismatches 184; Indels 66; Gaps 15;  
  
QY 37 DVFTAGSGPIGANYAKLCVE--AGLRVVMVEIGNAADSFYVNAEEGTAVYVPGYHKKE 94  
Db 33 DVFTAGSGPISCTYARHIDTSTKYMAEIGSQDN-----PVI-GAHHKS 79  
  
QY 95 IEFQKIDRVNVNIKALQOVSPVRNQNPTLDPGAWSPGSSA-----ISNGKNPHOR 150  
Db 80 IKQKIDKFVNIINGALQISPSPTYQTLAVAAW-APPIDPAGQLVIMGNPQOE 138  
  
QY 151 EFENLSAEATRGVGGMSTHTWCTSTRIHPMESLPQIGRPKLSNDPAEDDKWENLYSE 210  
Db 139 AGLNPGSAVTRVGGMATHWTACPTPHD-----EERVNPDV-DKQEFDALLER 187  
  
QY 211 AERLIGSTKEFDESIRHTLVLSLODAYKDRORIFRPLPLACHRLKNAPEYVHWSAEN 270  
Db 188 AKTLNVHSDQYDSDSIHQIVKETLQOTL-DASRGVTTPLGVERRTDNPIYVTWTDGAD- 245  
  
QY 271 LFHSIYNDDRQKKLFTLLTNHRCRLALTCGYEKKKIGAEVRNMLLAPRNPSSOLDYSIMA 330  
Db 246 ---TVLGDVPKSPREALVTETRTKLIVSETNPQVVAALLRNL---NTSN--DELVA 296  
  
QY 331 KVTYVLSAGTGNQOILYNSGSGIQTVPNRNDSLIPNLGRYITQPMAFQOIVLRQEFVDS 390  
Db 297 KSEVIACGAVCTPQILWNS-----NIRPY-----ALGRYLSEQSMFTFCQIVLKRGI 345  
  
QY 391 VRDDPYGLPMWKEVAQHIANKPTDALPIPRDPEPOVTTPTTEHPHWHQIHRDAFSYG 450  
Db 346 IATDPR-----FAKVEAHKKHDDVLPFPFHEPEPQVMIPYTSDFPWHVQVHRDAFSY 401  
  
QY 451 AVGPEDSVSRVIVDLRFMGATDPEANLLVF-----QNDVQDGYSMQPQTFRYRPST 501  
Db 402 DVGPKADPRVVVDLRFPGKSDIYEENRVTFGPNPKLREWEAGVTDYGMQPTFFHVKTN 461  
  
QY 502 ASNVARKMMADCEVASNIGGLYPTSPQPMFDPGLALHLAGTTRIGTDFKATTVADNNSL 561  
Db 462 ADGDRQDRQMMNDMTNVANMLGGYLPQSPQPMAPGLVHLITGTTRIGTDQTSVADPTSK 521  
  
QY 562 VWDFANLYVAGNTIRTFGENPTLTSMCHAIKSARSIIINTL 603  
Db 522 VHNFNLMWVGNGCIPDATAACNPTRTSVAVALKGAEEAVVYL 563  
  
RESULT 4  
US-09-023-731-12  
; Sequence 12, Application US/09023731

; Patent No. 6291648  
; GENERAL INFORMATION:  
; APPLICANT: Kawamura, Yukio; Morita,  
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.  
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND  
; TITLE OF INVENTION: GENE ENCODING SAME  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MICROSOFT WORD 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,731  
; FILING DATE: 13-FEB-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 29275/1997  
; FILING DATE: 13-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNETH H. SONNENFELD  
; REGISTRATION NUMBER: 33,285  
; REFERENCE/DOCKET NUMBER: 3479-4000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 51  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-023-731-12  
  
Query Match 5.8%; Score 192; DB 4; Length 51;  
Best Local Similarity 62.7%; Pred. No. 1e-12;  
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;  
  
QY 405 VAQHIANKPTDALPIPRDPEPOVTTPTTEHPHWHQIHRDAFSYGAVGPE 455  
Db 1 VEAHKKHPDDVLPFPFHEPEPQVMIPYTSDFPWHVQVHRDAFSYGDVGR 51  
  
RESULT 5  
US-09-297-937C-9  
; Sequence 9, Application US/09297937C  
; Patent No. 6337199  
; GENERAL INFORMATION:  
; APPLICANT: YUN, Do Young  
; APPLICANT: PAN, Jae Gu  
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence  
; TITLE OF INVENTION: Encoding the Same and Production of 2-keto-D-Gluconate  
; TITLE OF INVENTION: Using Transformed Recombinant E. Coli  
; FILE REFERENCE: P66159US0  
; CURRENT APPLICATION NUMBER: US/09/297,937C  
; CURRENT FILING DATE: 1999-05-11  
; PRIOR APPLICATION NUMBER: PCT/KR98/00296  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: KR 97-48802  
; PRIOR FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 615  
; TYPE: PRT

COMPUTER READABLE FORM:



```
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-023-731-11

Query Match 4.0%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 3.1e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

QY 333 YVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYIEQPMAPCOIVLROEFVDSVR 392
Db 2 FVIACGAVCTPQILWNS-----NIRPY-----ALGRYLSQSMTCQIVLKRGIYDAIA 50
QY 393 DDP 395
Db 51 TDP 53

RESULT 8
US-09-265-108-2
; Sequence 2, Application US/09265108
; Patent No. 6033891
; GENERAL INFORMATION:
; APPLICANT: Gollightly, Elizabeth
; APPLICANT: Brown, Kimberly
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850,000-US
; CURRENT APPLICATION NUMBER: US/09/265,108
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-265-108-2

Query Match 3.6%; Score 118.5; DB 3; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETYNDGVDFPIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYVNAEETAVP- 84
Db 242 APEETY-----DVIWVGAGAGGIPVADKLSEAGHKVLLIEKGP-----STGRMQGTMKPE 292
QY 85 YPCYHKKNEIEFOKIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSAAPPSSAISNG 144
Db 293 WLEG-----TDLTRF--DVPGLCNQIWDSAGIACITDQDAGCVLGGGTAVNAG 340

; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-023-731-11

Query Match 4.0%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 3.1e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

QY 333 YVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYIEQPMAPCOIVLROEFVDSVR 392
Db 2 FVIACGAVCTPQILWNS-----NIRPY-----ALGRYLSQSMTCQIVLKRGIYDAIA 50
QY 393 DDP 395
Db 51 TDP 53

RESULT 9
US-09-479-264-2
; Sequence 2, Application US/09479264
; Patent No. 6280976
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Gollightly
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850,200-US
; CURRENT APPLICATION NUMBER: US/09/479,264
; CURRENT FILING DATE: 2000-01-05
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-479-264-2

Query Match 3.6%; Score 118.5; DB 4; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETYNDGVDFPIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYVNAEETAVP- 84
Db 242 APEETY-----DVIWVGAGAGGIPVADKLSEAGHKVLLIEKGP-----STGRMQGTMKPE 292
QY 85 YPCYHKKNEIEFOKIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSAAPPSSAISNG 144
Db 293 WLEG-----TDLTRF--DVPGLCNQIWDSAGIACITDQDAGCVLGGGTAVNAG 340

QY 145 K--NPHOREFENLSAEAVTRGVGCMSTHTWCSTPRIHPPMESLPGIGRPKLSNDPAED-- 200
Db 341 LWMKPIDLWDENFPE-----GHSQDLAAATERV---FERIPGTWHPSDGKLYRDEG 391
QY 201 -----DKWNELYSEAEIRLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRFRPLPL 251
Db 392 YKVLSSGLAESGWKEV-----VANEVNPENKNTFAHTHFAGGE-----RNGPL 436
QY 252 ACHRLKNAPEYVEWHSANLFLHSIYNDKQKFLFTLTNRCTRLALTGGYKKGAAEV 311
Db 437 ATY-----LVSADAREN-----PSLATNTAVRAVRTGG---KVTGVEL 472
QY 312 RNLLATRNPSQLDSYIMAKYVVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYI 371
Db 473 ECL--TDGYSGIVKLNKGGVIFISAGAFSGAKLLFRSGIG----- 511
QY 372 TEOPMAFCQIVLR---QEFVDSVRDDPYGLPMWKEAVAOHIAKN-PTDAL---PIPER 422
Db 512 ---PEDQLRVVASSKOGEDFIDE-KD-----WIKLPVGINLIDHLNTDLILTHPDVVFY 561
QY 423 DPEQVTPTEEHPWHTQIHRDAFSYGA--VGP-----EVDSRVIVDL 464
Db 562 DFEAWTTPTEADKQLYLEQRSGLAQAPNIGPMWKEQVTPSDGITRQFQWTVARVEGDS 621
QY 465 RWFQATDPEANLLVFNQVQDQVGS-----MPQPTFRYRPSTASNVRAKMMADM 514
Db 622 RFTNSSHAMTLSQYLRGVVSRGRATITQGLVTTVAEHPYLHNAGDEAVIQGKLNIES 681
QY 515 CEVASNLGGVLPSTSP---QFMDPGL-----ALHLACTTRIGFDK---ATTVADN 558
Db 682 LNVIPNTWLP--PPGSTVEEYVDSLLVSASARRSNHMMGTAKLGTDDGRYGTSVVDL 739
QY 559 NSLWMDPANLYVAGNGTIRTFGE-NPTLASMCHAIKSARSIIITLK 604
Db 740 DTKVYGTDLNLFV-DASIFPGMSTGPNPSAMIVIAAEQAERILKLRK 785

RESULT 9
US-09-479-264-2
; Sequence 2, Application US/09479264
; Patent No. 6280976
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Gollightly
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850,200-US
; CURRENT APPLICATION NUMBER: US/09/479,264
; CURRENT FILING DATE: 2000-01-05
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-479-264-2

Query Match 3.6%; Score 118.5; DB 4; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETYNDGVDFPIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYVNAEETAVP- 84
Db 242 APEETY-----DVIWVGAGAGGIPVADKLSEAGHKVLLIEKGP-----STGRMQGTMKPE 292
QY 85 YPCYHKKNEIEFOKIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSAAPPSSAISNG 144
Db 293 WLEG-----TDLTRF--DVPGLCNQIWDSAGIACITDQDAGCVLGGGTAVNAG 340

QY 145 K--NPHOREFENLSAEAVTRGVGCMSTHTWCSTPRIHPPMESLPGIGRPKLSNDPAED-- 200
```

Db 341 LWWKPTDLWDENFPE-----GWSQDLAAATERV---FERIPGTWHPMSMDGKLYRDEG 391  
QY 201 -----DKWNNELYSAEALRGTSKEPDESIRHTLVLSLODAYKDRORIRPLPL 251  
Db 392 YKVLSSGLAESGWKE-----VANEVPEKRTFAHTFMFAGGE-----RNGPL 436  
QY 252 ACHRLKNAPEYVEHSAENLFSIYNDOKKKLTLLTNHRCRLALTGGEKKIGAAEV 311  
Db 437 ATY-----LVSADAREN-----FSLWNTAVRAVRIGG---KVTGVEL 472  
QY 312 RNLATRNPSOLDYSIMAKVYVLASGAIGNPQILYNSGSLGVTQPRNDSLIPNLGRYI 371  
Db 473 ECL--TDGYSIGVILNKGEGVIFSAGAFGSAKLLFRSGIG-----511  
QY 372 TEQPMACQIVLR---QEFVDSVRDDPYGLPWKEAQAHIKAKN-PTDAL-----PIPR 422  
Db 512 ----PEQURVASSKGDGFIDE-KD-----WIKLPVGYNLIDHLNLTDLILTHPDVVFY 561  
QY 423 DPEQVTPPTFEEHPWHTQIHRDAFSYGA--VGP-----EVDSRVIVDL 464  
Db 562 DFEAWTTPIEADKQILYLEQSRGILAQAPNIGPMWEGVTPSDGITRQFQWTVARVEGDS 621  
QY 465 RWFEGATDPEANLLVFQNDVQDYS-----MPOFTFRRPSTASNVRAKMMADM 514  
Db 622 RFTNSHAMTLSQLGRGVYVSRGRATITQGLVTVAEHPYHLNAGDKAEAVIQIKNLIES 681  
QY 515 CEVASNLGGYLPSTPP---QFMDPGL-----ALHLAGTTRIGFDK-----ATTVDN 558  
Db 682 LNVINITWVLP--PPGSTVEYVDSLLVSASARSNSHWMGTAKLIGTDGGRYGGTSVVDL 739  
QY 559 NSLVWDFANLYVAGNGTIRGTGE--NPILTSCHAIKASRSIINTLK 604  
Db 740 DTKVGTDLNLFVV-DASIFPGHSTGNPSAMIVIAEQAERILKLR 785

RESULT 10

US-09-023-731-14

; Sequence 14, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; APPLICANT: Akhihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; TITLE OF INVENTION: GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 29275/1997

; FILING DATE: 13-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: KENNETH H. SONNENFELD

; REGISTRATION NUMBER: 33,285

; REFERENCE/DOCKET NUMBER: 3479-4000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 50

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-023-731-14

Query Match 3.5%; Score 115; DB 4; Length 50;

Best Local Similarity 53.5%; Pred. No. 9.7e-05;

Matches 23; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 484 VODGYSMQPTFRYRSTASNVRAKMMADMCEVASNLGGYLP 526

Db 8 VTDTYGMQPTFHVRTNADGDRQDMNDMTNVANMLGGYLP 50

RESULT 11

US-09-023-731-15

; Sequence 15, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; APPLICANT: Akhihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; TITLE OF INVENTION: GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 29275/1997

; FILING DATE: 13-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: KENNETH H. SONNENFELD

; REGISTRATION NUMBER: 33,285

; REFERENCE/DOCKET NUMBER: 3479-4000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 35

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-023-731-15

Query Match 3.3%; Score 108; DB 4; Length 35;

Best Local Similarity 66.7%; Pred. No. 0.00028;

Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 528 SPPQPMDFGLALHLAGTTRIGFDKATTVADNNS 560

Db 2 SYPQPMAGPLVHLHTGTTRIGTDDQTSVADPTS 34

RESULT 12

US-08-609-049A-12

; Sequence 12, Application US/08609049A  
; Patent No. 5948664

## GENERAL INFORMATION:

; APPLICANT: Williams, Lewis T.  
; APPLICANT: Molz, Lisa  
; APPLICANT: Chen, Yen-Wen  
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/609,049A  
; FILING DATE: 29-FEB-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Karen B.  
; REGISTRATION NUMBER: 29,684  
; REFERENCE/DOCKET NUMBER: 2307K-063700US

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422

## INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1876 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-609-049A-12

Query Match 3.2%; Score 104.5; DB 2; Length 1876;  
Best Local Similarity 22.0%; Pred. No. 0.72;  
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;

QY 83 VPYPGVYHKKNEIEFQKIDRFVNVKIGALQVSVVPRNONVPTLDPGA---WSAP--P 136  
||| | : : : : : ||| | : : : : :  
DB 184 VPQPAQQQRPLN-SEELQRLYSM---PAQMAVVPVQPNAYMYPGAVVPTATIPV 239  
||| | : : : : : ||| | : : : : :  
QY 137 GSSAISNGKNPHOREFENLSAEAVTRGVGMSTHTWCTPRIHP-PMESLPGIG--RPKL 193  
||| | : : : : : ||| | : : : : :  
DB 240 GSAAFMPQYPAQ-----GYGFGAYTHMDLRRPQSPAPQOAPTATTSHHHSQP 288  
||| | : : : : : ||| | : : : : :  
QY 194 SNDAEDDKENNELYSEAEER---LIGTSKEFDESIRHTLVLSLQDA---YKDRQRI 245  
||| | : : : : : ||| | : : : : :  
DB 289 SNHSTSPAPANGVAFPARQVFTGVGSSSHTGNGGSHSVPRRNDLIDLNHEDYSRV 348  
||| | : : : : : ||| | : : : : :  
QY 246 -----FRPLPLACHRLKNAPEYVHWSAENLHFSIYNDKQKKLFTLLTNHRCRLALTG 300  
||| | : : : : : ||| | : : : : :  
DB 349 SVLEAFDPLNDNTGNTASDTSIYAEYDPDFLYSGD-----AATQYSDPMYEAVN 401  
||| | : : : : : ||| | : : : : :  
QY 301 GYEKKIGAAEVR-NLLATR-----NPSQLDSYIMA-----KYVILASGAIGNP----- 343  
||| | : : : : : ||| | : : : : :  
DB 402 RWDKTATVSPNVLGWRQDFLSQPSSTSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461  
||| | : : : : : ||| | : : : : :  
QY 344 --QILYNSGFGSLQVLT-PRNDSLIPNLGRYITQPMAFQCOIVLRQ-----EFVDSVR 392  
||| | : : : : : ||| | : : : : :  
DB 462 RNOOCYESNOAMPVSRPPOSSVLTD---SYTSSIP---ANVLDRRKTKTRLVELISDQR 516  
||| | : : : : : ||| | : : : : :  
QY 393 -DDPYGLPMWKEAQAQIAKNPTDALP 418  
||| | : : : : : ||| | : : : : :  
DB 517 TDDPELLEFY-HMWKEVRYRPHDDAP 542  
||| | : : : : : ||| | : : : : :

## RESULT 13

US-09-170-996-12  
; Sequence 12, Application US/09170996  
; Patent No. 6291220

## GENERAL INFORMATION:

; APPLICANT: Williams, Lewis T.  
; APPLICANT: Molz, Lisa  
; APPLICANT: Chen, Yen-Wen  
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/170,996  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/609,049  
; FILING DATE: 29-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Karen B.  
; REGISTRATION NUMBER: 29,684  
; REFERENCE/DOCKET NUMBER: 2307K-063700US

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422

## INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1876 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-170-996-12

Query Match 3.2%; Score 104.5; DB 4; Length 1876;  
Best Local Similarity 22.0%; Pred. No. 0.72;  
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;

QY 83 VPYPGVYHKKNEIEFQKIDRFVNVKIGALQVSVVPRNONVPTLDPGA---WSAP--P 136  
||| | : : : : : ||| | : : : : :  
DB 184 VPQPAQQQRPLN-SEELQRLYSM---PAQMAVVPVQPNAYMYPGAVVPTATIPV 239  
||| | : : : : : ||| | : : : : :  
QY 137 GSSAISNGKNPHOREFENLSAEAVTRGVGMSTHTWCTPRIHP-PMESLPGIG--RPKL 193  
||| | : : : : : ||| | : : : : :  
DB 240 GSAAFMPQYPAQ-----GYGFGAYTHMDLRRPQSPAPQOAPTATTSHHHSQP 288  
||| | : : : : : ||| | : : : : :  
QY 194 SNDAEDDKENNELYSEAEER---LIGTSKEFDESIRHTLVLSLQDA---YKDRQRI 245  
||| | : : : : : ||| | : : : : :  
DB 289 SNHSTSPAPANGVAFPARQVFTGVGSSSHTGNGGSHSVPRRNDLIDLNHEDYSRV 348  
||| | : : : : : ||| | : : : : :  
QY 246 -----FRPLPLACHRLKNAPEYVHWSAENLHFSIYNDKQKKLFTLLTNHRCRLALTG 300  
||| | : : : : : ||| | : : : : :  
DB 349 SVLEAFDPLNDNTGNTASDTSIYAEYDPDFLYSGD-----AATQYSDPMYEAVN 401  
||| | : : : : : ||| | : : : : :  
QY 301 GYEKKIGAAEVR-NLLATR-----NPSQLDSYIMA-----KYVILASGAIGNP----- 343  
||| | : : : : : ||| | : : : : :  
DB 402 RWDKTATVSPNVLGWRQDFLSQPSSTSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461  
||| | : : : : : ||| | : : : : :  
QY 344 --QILYNSGFGSLQVLT-PRNDSLIPNLGRYITQPMAFQCOIVLRQ-----EFVDSVR 392  
||| | : : : : : ||| | : : : : :

Db 462 RNQCYESNQAAMPYRPPQSVLTD--SYTSSIP---ANVLDRRKTCRTLRYELISDQR 516

Qy 393 -DDPYGLPWKKAQAQIAKNPDTALP 418

Db 517 TDDPELLEFY-HWKEVRARYPHDDAP 542

## RESULT 14

US-09-023-731-6

; Sequence 6, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawanura, Yukio; Morita,

; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023.731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 29275/1997

; FILING DATE: 13-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: KENNETH H. SONNENFELD

; REGISTRATION NUMBER: 33,285

; REFERENCE/DOCKET NUMBER: 3479-4000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 61

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-023-731-6

Query Match 3.1%; Score 102.5; DB 4; Length 61;

Best Local Similarity 43.5%; Pred. No. 0.0027;

Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

Qy 104 FVNVKGLQOVSVVRNQNVTLPDPAWSAPPGSSA-----ISNGKNPHQREFNLSAE 159

Db 1 FVNIINGALQPISSPSTYQPTLAVAAW-APPIDPAGCOLVINGHNPQEAAGLNLP 59

Qy 160 VT 161

Db 60 VT 61

## RESULT 15

US-08-549-515-5

; Sequence 5, Application US/08549515

; Patent No. 6054123

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: Haemophilus Influenzae

; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 Unviersity Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/549,515

; FILING DATE: 27-OCT-1995

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-522

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1153

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 806 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-549-515-5

Query Match 3.1%; Score 100.5; DB 3; Length 806;

Best Local Similarity 18.6%; Pred. No. 0.42;

Matches 69; Conservative 56; Mismatches 139; Indels 107; Gaps 17;

Qy 31 YGNGVDVFIAGSGPIGATYAKLQVE----- 56

Db 230 FGNPAETRMGG---GLTY---CIEQAKARSNAKMIIDPRYNDTGAGREDEWIPRPG 283

Qy 57 -----AGLRVVMVEIGAAD----SFYAVNAEETGTAVPYVP--GYHKKNEIEFQKIDIRF 104

Db 284 TDAALVAALAVYMIQENLVDPFLDKYCVGYDEKTLPADAPKNGHYKAVILGYGN--DGI 341

Qy 105 VNVKGLQOVSVVRNQNVTLPDPAWSAPPGSSAISNGKNPHQREFNLSAEAVTRGV 164

Db 342 AKTPEWAAKITGIPA--ERIIKLAREIGSTKP--AFISQGWGPQRRS-----NGELISR 393

Qy 165 -----GGMSTHTWCSTPRTHPPMESLPGIGRPKLSNDPAEDDKENNELYSEAERLIGT 217

Db 394 AMLPILTNVGIHGGNTGAR--ESAYSIPFVPMPTLK-----NPMKASIPMFLGT 441

Qy 218 STKEPDESIRHTLVLRSLQADYKQRIERFPLPLACHRLKNAPEYVEVHSAENLFHSYN 277

Db 442 -----DAIIRGT-EMTALTDGIRGVDKLSPPIKVWNVYASNC--LINQHAQINRTHDILQ 493

Qy 278 DDQKKKFTLLTNHRCTRIALTGGYEKKIGAAEVNLLATRNPSQ-----LDSTYIMAKV 332

Db 494 DDTQCEMIITIDNHMTS-----TAKYSDILLPDCPTSEQMDFDALDAFVSNMA 540

Qy 333 YVLASGAIGNP 343

Db 541 YVIFADQVIKP 551

Search completed: September 3, 2002, 15:57:29

Job time: 3139 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 14:57:41 ; Search time 82.18 Seconds  
(without alignments)  
835.284 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSFEQMLRDPMSQING.....INTLKGTDGKNTGEHRL 618

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3284	100.0	618	22	Lyophyllum shimeji
2	1232.5	37.5	623	20	Coriolus versicolor
3	1232	37.5	633	21	Pleurotus cornucop
4	1218.5	37.1	623	17	Pyranose oxidase.
5	1210	36.8	622	22	Trametes hirsuta p
6	1091	33.2	566	20	An antitumour prot
7	1044	31.8	564	21	T. matsutake pyran
8	1044	31.8	564	21	Trichoderma deri
9	200	6.1	544	20	G. oxydans D-sorbi
10	192	5.8	51	20	Peptide derived fr
11	191.5	5.8	615	20	Gluconate dehydro

12	177	5.4	67	21	AA10459	T. matsutake pyran
13	177	5.4	67	21	AA10459	N-terminal fragmen
14	168	5.1	34	22	AA10459	Lyophyllum shimeji
15	162	4.9	30	22	AA10459	Lyophyllum shimeji
16	159.5	4.9	657	21	AA10459	Arabidopsis thalia
17	159.5	4.9	675	21	AA10459	Arabidopsis thalia
18	159.5	4.9	748	21	AA10459	Arabidopsis thalia
19	142	4.3	29	22	AA10459	Arabidopsis thalia
20	130.5	4.0	58	20	AA10459	Peptide derived fr
21	121	3.7	24	22	AA10459	Lyophyllum shimeji
22	121	3.7	704	20	AA10459	Candida tropicalis
23	120.5	3.7	546	20	AA10459	Choline oxidase (C
24	119	3.6	500	22	AA10459	A. tumefaciens 1,5
25	119	3.6	545	21	AA10459	Arabidopsis thalia
26	119	3.6	727	21	AA10459	Arabidopsis thalia
27	119	3.6	746	21	AA10459	Arabidopsis thalia
28	119	3.6	2639	22	AA10459	Novel human diagno
29	118.5	3.6	785	21	AA10459	Humicola insolens
30	118.5	3.6	785	22	AA10459	H. insolens DSM 18
31	116.5	3.5	698	20	AA10459	Candida cloacae fa
32	115	3.5	50	20	AA10459	Peptide derived fr
33	110	3.3	551	21	AA10459	Streptomyces globi
34	108	3.3	35	20	AA10459	Peptide derived fr
35	107.5	3.3	964	22	AA10459	Amino acid sequenc
36	106.5	3.2	778	20	AA10459	Chlamydia pneumoni
37	106	3.2	572	22	AA10459	S. epidermidis ope
38	106	3.2	1356	22	AA10459	Drosophila melanog
39	105	3.2	609	22	AA10459	Propionibacterium
40	105	3.2	833	22	AA10459	Drosophila melanog
41	104.5	3.2	1876	18	AA10459	Phosphatidyl inosi
42	104.5	3.2	1876	22	AA10459	Drosophila melanog
43	104.5	3.2	1876	22	AA10459	Drosophila melanog
44	103	3.1	437	22	AA10459	Novel human diagno
45	103	3.1	1686	19	AA10459	Human class II P13

#### ALIGNMENTS

#### RESULT 1

AA10459

ID: AA10459 standard; Protein: 618 AA.

XX AA10459;

AC AA10459;

XX AA10459;

DT 20-JUL-2001 (first entry)

XX Lyophyllum shimeji antibacterial protein.

DE Lyophyllum shimeji antibacterial protein.

KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;

KW Pyricularia orizae; Rhizoctonia solani; rice pathogen.

XX Lyophyllum shimeji.

OS Lyophyllum shimeji.

XX WO200121657-A1.

PN 29-MAR-2001.

XX 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NISB) JAPAN TOBACCO INC.

XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI; 2001-281598/29.

XX N-PSDB; AAF99980.

XX Antibacterial protein and encoded gene isolated from Lyophyllum

PT shimeji, with activity against plant pathogenic bacteria, applicable in

PT agriculture e.g. rice cultivation at low concentration, produced at low

PT cost on large scale -

PS Claim 3; Page 43-45; 52pp; Japanese.

CC The present sequence is an antibacterial protein from the fungus  
 CC *Lyophyllum shimeji*. The protein was obtained from a fraction prepared  
 CC by extracting *Lyophyllum shimeji* with water and subjecting the extract  
 CC to ammonium sulphate precipitation. The protein inhibits the growth  
 CC of the plant pathogenic bacteria *Pseudomonas* and *Rhizoctonia*  
 CC *solanii* at a relatively low concentration. *P. oryzae* and *R. solani* are  
 CC causative of the two major diseases of rice. The protein contains  
 CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.  
 CC The antibacterial protein can be produced at low cost on a large scale.

XX Sequence 618 AA;

Query Match 100.08; Score 3284; DB 22; Length 618;

Best Local Similarity 100.08; Pred. No. 2.1e-299;  
 Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSTEQMLRDYPRSGMQINGQIPKNAIHEITYGNDGVDFIAGSGPIGAYAKLCVEAGLR 60

Db 1 mslsteqmlrdyprsmqngqipknaihetygndgvdfiagsgppgatyaklcveaglr 60

QY 61 VVMVEIGAADSFYAVNAEGTAVPYVPGYHKKNEIEFFQKIDRFVNVIKGALQQVSPVVR 120

Db 61 vvmveigaadsfyavnaegstavpyvpgyhkkneiefqkldrfvnvirkalqgvsvpr 120

QY 121 NONVTLDPGANSAPGSSAISNGKNPHOREFENISAEAVTRGVGGMSTHTWCSTPRIHP 180

Db 121 nqnvptldpgavapgsaisangknphorefenisaeavtrgvvgmsthtwcstprihp 180

QY 181 PMSLPGIGRPKLSNDPAEDKEMNELSEAEERLIGTSTKPEDESIRHTLVLSRLQDAYK 240

Db 181 pmslpgigrpklsndpaedkewnelseaeerlgtstkefdesirhtlvlsrlqdayk 240

QY 241 DRQIFRPLPLACHRLKNAPEYVHSAENLPHSYNDKQKFLTLTNHRCRLALTG 300

Db 241 drqifrpplachrlknapeyvhhsaenlfhsynddkqkfltltnhrcrlaltg 300

QY 301 GYEKIGAEVNNLLATRNPSQLDSYIMAKYVVLASGAINQPIILYNSGFGSLQVTPRN 360

Db 301 gyeckigaeavnllatrnpsqldsyimakvvyilasgaignpqillynsqfsglqvtrpn 360

QY 361 DSLIPNLGRYITEQPMAFQCIIVLROEFVDSVRDDPYGLPWWKEAVAQHIKAKNPTDALPIP 420

Db 361 dsllpnlgryteqpmafqciivlrgefvdsvrddpyglpwkavaqhiaknptdalpip 420

QY 421 FRDPEQVTPPTTEHPWHQTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANNLLVF 480

Db 421 frdpeqvtpptteehpwhqtqihrdafsygavgpevdsrvivdlrwfgatdpeannllvf 480

QY 481 QNDVQGYFMPQPTRYRPRSTASNVARKMADMCVEASNLGGYLPSPQPMDFGLALH 540

Db 481 qndvqgyfmpqptfryrprstasnvrarkmadmcveasnlggylpsppqfmdpglalh 540

QY 541 LAGTTRIGFDKATTVADNNLSVDFANLYVAGNGTIRTFEGENPTLTSMCHAIKARSII 600

Db 541 lagttrigfdkattvadnnlsvdfanlyvagngtirtfegenptltsmchaikarsii 600

QY 601 NTLKGGTGKNTGEHRL 618

Db 601 ntlkgtgdkntgehrnl 618

RESULT 2

AAW94308

ID AAW94308 standard; protein: 623 AA.

XX

AC AAW94308;

XX

DT 08-APR-1999 (first entry)

XX

DE Coriolus versicolor pyranose oxidase.

XX

KW Coriolus versicolor; modified: pyranose oxidase; heat stability;  
 KW pH stability; glucose; glucosone.

XX

OS Coriolus versicolor.

XX

PN JP11009271-A.

XX

PD 19-JAN-1999.

XX

PF 24-JUN-1997; 97JP-0181865.

XX

PR 24-JUN-1997; 97JP-0181865.

XX

PA (KIKK ) KIKKOMAN CORP.

XX

DR WPI; 1999-145891/13.

XX

PT New modified pyranose oxidase gene and protein - useful for  
 PT efficient recombinant production of pyranose oxidase with excellent  
 PT Km, heat and pH stability

XX

PS Claim 1; Page 7-9; 10pp; Japanese.

XX

CC The present sequence represents pyranose oxidase derived from *Coriolus*  
 CC versicolor, which can be modified by adding, deleting or replacing at  
 CC least one amino acid (preferably where the amino acid at position 542  
 CC is replaced) and still retains pyranose oxidase activity. The modified  
 CC pyranose oxidase with the amino acid at position 542 replaced, has the  
 CC following properties: (a) Action: It oxidises glucose to glucosone;  
 CC (b) Substrate specificity: It reacts specifically with glucose and also  
 CC reacts with galactose, L-sorbose, D-xylose and 1,5-anhydro-D-glucitol;  
 CC (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)  
 CC Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;  
 CC and (f) Temperature stability: Stable up to 55 degrees Celsius. The  
 CC present invention also describes a method for the preparation of a  
 CC modified pyranose oxidase. The method efficiently prepares a modified  
 CC pyranose oxidase with excellent Km value, heat stability and  
 CC pH stability.

XX Sequence 623 AA;

SQ

Query Match 37.5%; Score 1232.5; DB 20; Length 623;

Best Local Similarity 45.0%; Pred. No. 1.7e-106;

Matches 273; Conservative 73; Mismatches 176; Indels 85; Gaps 16;

QY 37 DVFIAAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEGTAVPYVPGYHKKNEIE 96

Db 48 dvvlgspigctyarelveagykvmfdgldgldki-----gahkktve 95

QY 97 FKQIDRFVNVTKGALQQVSVPRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENUS 156

Db 96 yqknldkfnnvlgqqlmsvvpvntlvtdltsptswqa--ssffvrngnpeqdlrnl 153

QY 157 ABATVTRGVGGMSTHTWCSTPRIHPMWSLPGIGRPKLSNDPAE-DDKEMNELSEAEERLI 215

Db 154 gqavtrvggmsthtwcstprfhdreq-----rpllvkddqdaadaewdrlytkaesyf 206

QY 216 GTSTKEFDESIRHTLVLSRLQDAYKDRQIFRPLPLACHRLKNAPEYVHSAENLH--H 273

Db 207 ktgtqgkiesirhnvlvnlklaeeykg-qrdfqgiplaatr--rsptfveasantvfdlq 263

QY 274 SIYNDKQKFLTLTNHRCRLALTGGEYKKGKIGAEVNNLLATRNPSQLDSY----- 327

Db 264 nrpntdapnerfnlfpavacerv-----vrn-----tsnseieslhlhdl 304

QY 328 -----IMAKYVVLASGAINQPIILYNSGFGSL-QVTPRN-DSLIPNLGRYITEQPMAPC 379

Db 305 egdrfeikadvilvtagavhnaqllvnsqfqlgrpdpnanpqlpilsyisgteivfc 364

```

QY 380 QIVLRQEFVDSVRDDPY--GLP-----WWKEAVAQHIKAKNPFDAL 417
Db 365 qtmstmlidsvksdmiirgnpgdglvsyvtypgaetknkhpwwnekvknhmmqhgedp1 424
QY 418 PIPRPDPEQVTPPTTEHEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFQATDPEANNL 477
Db 425 pifedpepavtllfqshpwhqihrdafsygavqsidsrlivdwrfgrtepkkeen 484
QY 478 LVFQNDVQDGYSPQPTFRYR--PSTASNVARKMMADCEVASNLGGYLYLTPSPQFMDPG 536
Db 485 lwfsdkitdtnmpqptfdrfpaqrtskeadmtdmcmvmsakigglpgslpqfmeqg 544
QY 537 LAHLACTTRIGDKA--TTVADNLSLVDFANLYVAGNCTIRTFGPGENTLFSMCHAIK 594
Db 545 lvhlqgthrmgfdgedekccvntdsrvfgknlfgcgcnptayganptltamslaik 604
QY 595 SARSIIN 601
Db 605 sceyikn 611

RESULT 3
AAV52700
ID AAY52700 standard; Protein; 633 AA.
XX
AC AAY52700;
XX
DT 07-MAR-2000 (first entry)
XX
DE Pleurotus cornuopiae antitumour protein.
XX
KW Antitumour; cancer; tumour; treatment; expression; tumour suppressor;
KW p53; pBR; ss.
XX
OS Pleurotus cornuopiae.
XX
PN JP11315096-A.
XX
PD 16-NOV-1999.
XX
PF 07-AUG-1998; 98JP-0236349.
XX
PR 08-AUG-1997; 97JP-0215311.
PR 02-MAR-1998; 98JP-0066176.
XX
PA (NEW-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.
XX
DR WPI; 2000-058170/05.
DR N-PSDB; AA246411.
XX
PT An antitumour protein derived from Pleurotus cornuopiae and its gene -
PT useful for treatment of cancer including those caused by abnormal
PT expression of cancer inhibitory gene (e.g. p53 and pBR)
XX
PS Claim 3; Page 13-15; 23pp; Japanese.
XX
CC The invention relates to a novel antitumour protein extracted from
CC fruiting bodies of the fungus Pleurotus cornuopiae. The protein and
CC nucleotides encoding it are useful for the treatment of cancer,
CC including those caused by abnormal expression of tumour suppressor
CC genes such as p53 and pBR. This sequence represents the antitumour
CC protein.
XX
SQ Sequence 633 AA;

Query Match 37.5%; Score 1232; DB 21; Length 633;*
Best Local Similarity 44.6%; Prod. No. 2e-106;
Matches 275; Conservative 77; Mismatches 193; Indels 72; Gaps 18;

QY 18 INQIIPKNAHEHYGNDGVDFYTAGSGPIGATYAKLCVEAGLRVWVVEIGAADSFYAVNA 77
Db 44 ipdklpsg-----dtskfdvrvvgspvgstyarlliveagfkvgmfeigeids----- 91

```

```

QY 78 EEGTAVYVPVGYHKKNIEFQKOIDREVNVIKQALQVSVPRNQNVPTLDPGAWSPGP 137
Db 92 --grkl-----gshkntveyknlidkfvhviqgqlmpsvpnkyvadtispaswqa--s 143
QY 138 SSAISNGKNPHORFENLSAEANTRGVGGMSTHTCTSPRIHPPMESLPGIGRPKL-SND 196
Db 144 thvrnganpeqpftnlsggavtrvvvggmthwcatprfhks-----erpklvkdd 196
QY 197 PAEDDKENNELYSABERLIGTSTKEFDESIRHTLVLSRLQDAYKDRIRFRLPLACHRL 256
Db 197 dsadatweryldiaesfvktghnqfcdgsirhnlvleklgesyag-grgfeqiplaaqrt 255
QY 257 KNAPEYVEHWSAENLF--HSIYNDKQKKLFTLLTNHRCRLALTGGYKKGIAAEVRNL 314
Db 256 n--prfvevssahvtfdlenrpnaddegrfnlfpavvcervt-rdsldrkievhdh 312
QY 315 LATRNPSSQLDSY-IMAKVYVVLASGAIGNPQILYNSFGSLQVTPRND-----SLIPNLGR 369
Db 313 lsg-----drykvkadvfildcsgavhmpqilvnsgr-grmgqpsdslppptlilpylgs 364
QY 370 YITEQPMACQIVLRQEFVDSVRDDP--YGLP-----WWKEAVAO 407
Db 365 yiteqtlitfcqvfstelnlvksdmliivgtppgdydvstfcdpsnkhpnwnnekvqk 424
QY 408 HIAKNPTDALPIPRDPEQVTPPTTEHEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWF 467
Db 425 hnmqhqedp1p1ddpepavtllfqshpwhqihrdafsygavaesidsrlvvdwrf 484
QY 468 GATDPEANLLVPONDVQDGYSPQPTFRYR--PSTASNVARKMMADCEVASNLGGYLP 526
Db 485 grtepvveenklwfskqitdaynlpqptfsfrfpgqrtaqeaelmmadmctmstkvggflp 544
QY 527 TSPQFMDPGLALHLAGTTRIGRFDKATVA--DNNSLVDFANLYVAGNCTIRTFGCGENP 584
Db 545 gsyppqfmapglvhlhgghrmgfddeadkacvdcnkvfmgmenifggcgngitayasn 604
QY 585 TLTSMSCHAIKSARSIIN 601
Db 605 tltavalairackeyirn 621

RESULT 4
AAV99628
ID AAR99628 standard; Protein; 623 AA.
XX
AC AAR99628;
XX
DT 27-NOV-1996 (first entry)
XX
DE Pyranose oxidase.
XX
KW Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
KW marker; diagnosis; 1,5-anhydro-D-sorbitol.
XX
OS Corioliolus versicolor.
XX
PN DE19545780-A1.
XX
PD 13-JUN-1996.
XX
PF 07-DEC-1995; 95DE-1045780.
XX
PR 24-MAY-1995; 95JP-0124835.
PR 07-DEC-1994; 94JP-0304086.
XX
(KIKK ) KIKKOMAN CORP.
XX
PA Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
PI Suzuki M;
XX
DR WPI; 1996-278990/29.
DR N-PSDB; AAT34420.

```

XX DNA encoding protein with pyranose oxidase activity at neutral pH -  
PT useful for the determination of glucose in body fluids or foods, or  
PT 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis  
XX  
PS Claim 1; Page 14-16; 22pp; German.  
XX  
CC The present sequence encodes a protein isolated from *Corioliolus versicolor*,  
CC which has the enzyme activity of pyranose oxidase (PO). The PO oxidises  
CC glucose to gluconone and has an optimum pH of 7-7.5. It has a mol. wt. of  
CC 290000 (determined by gel filtration) and is stable at around 50deg.C.  
CC PO can be used for measurement of glucose in, e.g. foods or body fluids,  
CC or 1,5-anhydro-D-sorbitol which is an important marker used in the  
CC diagnosis of diabetes.  
XX  
XX Sequence 623 AA;

Query Match 37.1%; Score 1218.5; DB 17; Length 623;  
Best Local Similarity 44.8%; Pred. No. 3.5e-105;  
Matches 272; Conservative 72; Mismatches 178; Indels 85; Gaps 16;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVPGYHKKNEIE 96  
Db 48 dvlvvgspigctyareiveagkykvmfdigdsglki-----gahkknkte 95

QY 97 FOKDIDRFVNVIKGALQOVSPVRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156  
Db 96 yqkndkfvnvgqqlmsvpyvntlvdtlspwqg--ssffvrngsnpeqdlrnl 153

QY 157 ABATRVGGGSMTHWCTSTPRIHPHMSLPGIGRPKLNDPAE-DDKENNELYSAEERLI 215  
Db 154 gqavtrvggmsthwtcatprfdeq-----rpllvkdqddaddaewdrlytkaesyf 206

QY 216 GSTKEFDESIRHTLVLSLQDAYKDRIRFPLPLACHRLKNAPEYVHWSAENLF--H 273  
Db 207 krtcdqfkesirhnlvlkkaeeykg-grdfqqlplaatr--rsptfvevssantvldlq 263

QY 274 SIYNDKOKKLTLLTNHRCRTRLATLGGYEKKIGAAAEVRNLLATRNPSQLDSY----- 327  
Db 264 nrpntdapnerfnlfpavacerv-----vrn-----tsnseieslhihdl 304

QY 328 -----INAKVYVLASGAIGNPQILYNSGFSL-QVTPRN-DSLIPNLGRVITEQPMATC 379  
Db 305 sgdrfeikadvfvtagavhnaqlvnsfgqglgrdpbanppqlipsrlytedgslvfc 364

QY 380 QIVLRQEFVDSVRDDPY--GLP-----WKEAVAQHTAKNPTDAL 417  
Db 365 qtvmtelidskdmilrgnpgdlgyvtytpgaetkhkdpdwnekvnmmqhqedpl 424

QY 418 PIPFRDPEQVPTPTTESEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANNL 477  
Db 425 pipfedpeqvvtlfpqshpwhqtghrdafsygavqqlidsrllvdrwffgrtepkcenk 484

QY 478 LVFQNDVDGSMQPTPRYR-PSTASVNRARKMADCEVASNLGGYLPSPPOFMDPG 536  
Db 485 lwfsdkitdtnmpqptdfdfpagrtskaedmmtdmcmvkwakigglfpgslpqlfmeqpg 544

QY 537 LALHLAGTRIGFDKA--TTVADNNLSVDFANLYVACNGTIRTGFGENPTLTSCHAIK 594  
Db 545 lvllhggthrmgfdedeqdekccevttdsrvfgniflgcgnlptayganpiltamslaik 604

QY 595 SANSIN 601  
Db 605 sceyln 611

RESULT 5  
AAB48832  
ID AAB48832 standard; Protein; 622 AA.  
XX  
AC AAB48832;  
XX

DT 13-MAR-2001 (first entry)  
XX  
DE Trametes hirsuta pyranose oxidase.  
XX  
KW Pyranose oxidase; expression construct; recombinant production;  
KW monosaccharide oxidation; 2-keto derivative;  
KW hydrogen peroxide production.  
XX  
OS Trametes hirsuta.  
XX  
PN US6146865-A.  
XX  
PD 14-NOV-2000.  
XX  
XX 05-MAY-1999; 99US-0305381.  
XX  
PR 08-JUN-1998; 98DK-0000774.  
PR 10-JUN-1998; 98US-0088724.  
XX  
FA (NOVO ) NOVO NORDISK AS.  
XX  
PI Schneider P, Christensen S, Lassen SF;  
XX  
DR WPI; 2001-049055/06.  
DR N-PSDB; AAC87518, AAC87519.  
XX  
XX Novel nucleic acid molecule encoding polypeptide having pyranose  
PT oxidase activity used to design oligonucleotide probes to identify and  
PT clone DNA encoding the polypeptide from different genera or species -  
XX  
PS Claim 5; Fig 1; 20pp; English.  
XX  
CC The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which  
CC encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also  
CC relates to expression constructs, expression vectors and recombinant  
CC cells comprising pyranose oxidase nucleic acid sequences, and the  
CC recombinant production of Trametes hirsuta pyranose oxidase. Pyranose  
CC oxidase catalyses the oxidation of several monosaccharides in the  
CC pyranose form at position C2 to produce 2-keto derivatives with the  
CC release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta  
CC pyranose oxidase may be used to produce the enzyme and to design  
CC oligonucleotide probes to identify and clone genomic pyranose oxidase  
CC cDNA or genomic DNA from different genera or species of microorganisms  
CC (fungi or bacteria). The present sequence represents pyranose oxidase  
CC from the fungus Trametes hirsuta.  
XX  
SQ Sequence 622 AA;

Query Match 36.8%; Score 1210; DB 22; Length 622;  
Best Local Similarity 44.6%; Pred. No. 2.2e-104;  
Matches 274; Conservative 71; Mismatches 197; Indels 72; Gaps 16;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVPGYHKKNEIE 96  
Db 48 dvlvvgspigctyareiveagkykvmfdigdsglki-----gshkknkte 95

QY 97 FOKDIDRFVNVIKGALQOVSPVRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156  
Db 96 yqkndkfvnvgqqlmsvpyvntlvdtlspwqg--stffvrngsnpeqdlrnl 153

QY 157 ABATRVGGGSMTHWCTSTPRIHPHMSLPGIGRPKL-SNDPAEDDKENNELYSAEERLI 215  
Db 154 gqavtrvggmsthwtcatprfek-----lqrpvlvkdskaddaewdrlytkaesyf 206

QY 216 GSTKEFDESIRHTLVLSLQDAYKDRIRFPLPLACHRLKNAPEYVHWSAENLF--H 273  
Db 207 krtcdqfkesirhnlvlkkaeeykg-vrdfqqlplaatr--qsptfvevssantvldle 263

QY 274 SIYNDKOKKLTLLTNHRCRTRLATLGGYEKKIGAAAEVRNLLATRNPSQLDSYIMAKVY 333  
Db 264 nrpndkqkfnlfpavacnvrndnanselv9-lvrdlhggkslt-----lkakvy 316



Qy	334	VLASGATGNQILYNSGFGSLQVTPRND-----SLIPNLGRVITSQPMAFQCOIVLRQEF 387
Db	317	iltagavhnaqlaaagfgql---grdpakpipsllpylgthitegtlvfcqvmstel 373
Qy	388	VDSVRDDP--VGLP-----WKEAVAQHIAKNPDTALPIPRDPE 425
Db	374	insvtadmtivgpgphdysvtytpgnpnknkhpdwnekvkkmhmdhqedpfpifedpe 433
Qy	426	PQVTPPTFEHPMHTQIHRDAFSYGAVPEVDSRVIVDLRWFGATDPEANNLLVFNQDYO 485
Db	434	pqvttlfgatpwhqtqhrdafsygavqgsidsrlivdwrffgrtepkkeenklwfsdkit 493
Qy	486	DGYSMPQPTFRYRSTASNVARKEMADACEVASNLGGYLPTSPFPQFMDPGLALHLAGT 545
Db	494	daynairptfdfr--fpggreaedmtdmcvmsakiggfipgssypqimepgvivhlggth 551
Qy	546	RIGFDKAT--TVADNNLSLVWDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSINTL 603
Db	552	rmgfdekadkccvtdsrvgfknflggcgnlptayaanptitamslaiksceyl---k 608
Qy	604	KGGTDGKNTGEHRN 617
Db	609	knfepsnpvkvkhn 622
RESULT 6		
AAW87531		
ID	AAW87531 standard; Protein; 566 AA.	
AC	AAW87531;	
DT	26-FEB-1999 (first entry)	
DE	An antitumour protein.	
KW	Antitumour protein; Tricholoma matsutake.	
OS	Tricholoma matsutake.	
PN	JPL0313876-A.	
PD	02-DEC-1998.	
PF	13-FEB-1998; 98JP-0031452.	
PR	13-FEB-1997; 97JP-0029275.	
PA	(MOMO-) MOMOYA KK.	
PA	(NORQ) NORINSUISANSO SHOKUHN SOGO.	
DR	WPI; 1999-074153/07.	
DR	N-PSDB; AAV83626.	
PT	An anti-tumour protein - prepared by culture of host cell	
PT	transformed by vector containing base coding sequence	
PS	Claim 1; Page 7-8; 15pp; Japanese.	
CC	The present sequence encodes an antitumour protein, and is isolated	
CC	from Tricholoma matsutake.	
SQ	Sequence 566 AA;	

```

Query Match      33.28; Score 1091; DB 20; Length 566;
Best Local Similarity 42.66; Pred. No. 2.9e-93;
Matches 248; Conservative 84; Mismatches 184; Indels 66; Gaps 15;

QY 37 DVFTAGSGPTGATYKACLCVE--AGLRVVMVIRIGAADSFYAVNAEGTAVPVPGYHKHNE 94
|||||  ||||  : : : ||| : | : | : | : | : | : | : | : | : | : | : | : |
Db 33 dvfiagsgpiscytrahidntstktymaegisqdn-----pvi-gahhkn 79

QY 95 IEFKRDIDRFVNVYIKGALQQVSVPRNVONVPTLDPGAWSAAPGSSA-----ISNCKNPHOR 150

```

Db	80	ikfkdkidkfvniingalqpsispsdyqtlavaav-appidpsegqlvimghnpqnc	138
Qy	151	EFENLSAEAVTRGVGGMSTHTWCSTPRIHPMPESLPCIGRPKLSNDPAEDKREWNELTSE	210
Db	139	aglnpssavtrtvggmahwtcacplphd-----eervnnpv-dkqefdalier	187
Qy	211	AERLIGTSTKEFDESIRHTLVLSRLQDAYKDRDIRFRPLACHRLKNAPEYVEVHSAEN	270
Db	198	aktllnvhsdqyddsirqivvketlqql-dasrgvttlplgvertdnpiyywtgad-	245
Qy	271	LFHSIYNDQKOKLFTLLTNHRCRRLALTGGYEKKIGAAEVRNMLLATRNPPSOLDYSYIMA	330
Db	246	---tvlgdvkpsprfalvtetrvtkliivsetnptqvvaallrnl----ntsn--delvva	296
Qy	331	KVYVLASGAIGNPOILYNSGFSGLQVTPRNDLSLIPNLGRVITPQMAFCQIVLROEFVDS	390
Db	297	kswiaagvactpqlwns-----nirpy-----algrylseqsntfcqvlkrgivda	345
Qy	391	VRDDPYGLPWWKEAVAQOHIAKRNPTDALIPFRDPEQVTTPTTEEHPWHTQIHRDAFSYG	450
Db	346	iatdpr---faakveahkkkhpddvlpipfhepeqvmpiytsdfpwhqvhrdafsyg	401
Qy	451	AVGPEVDSRVIVDLRWFAGTDPENALLVF-----QNDVQDGYSMQPQTFYRFPST	501
Db	402	dvgpkadprvvvdrlrfgkadiveenrvtfgnpklirewagvtdtygmppqtfhvkrtcn	461
Qy	502	ASNVYARKMADNCEVASNLGGYLPSPPOFMDPGLALHLAGTTRTIGFDKATTVADNNSL	561
Db	462	adgdrdrdmndmncvnamlgyipdsgypqfmapgvlvlhtgtrtigrtdqdtsvadptsk	521
Qy	562	VWDFEANYLVAGNGIRTGFGENPTLTSMCHAIKSARSIINTL	603
Db	522	vhnfnlwgngncipdatacnprrtsvayalkgaeavvnyl	563
RESULT	7		
AA	AAB10457		
ID	AAB10457	standard; Protein; 564 AA.	
XX	AA	AAB10457;	
AC	AA	AAB10457;	
DT	DT	11-DEC-2000 (first entry)	
DE	DE	T. matsutake pyranose oxidase protein.	
KW	KW	Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism	
KW	KW	diabetes mellitus.	
OS	OS	Tricholoma matsutake.	
PN	PN	JP2000175698-A.	
PD	PD	27-JUN-2000.	
PF	PF	16-DEC-1998; 98JP-0357423.	
PR	PR	16-DEC-1998; 98JP-0357423.	
PA	PA	(NISB ) JAPAN TOBACCO INC.	
DR	DR	WPI: 2000-478485/42.	
DR	DR	N-PSDB; AAA71487.	
XX	XX		
PT	PT	A reagent containing pyranose oxidase for the determination of pyranose	
PT	PT		
PS	PS	Claim 5; Page 17-18; 23pp; Japanese.	
XX	XX		
CC	CC	This invention describes a novel reagent containing pyranose oxidase if	
CC	CC	the determination of pyranose which can be prepared from a fraction	
CC	CC	precipitated from an aqueous extract of matsutake mushroom by ammonium	
CC	CC	sulfate precipitation and has an antibacterial activity against at least	

A reagent containing pyranose oxidase for the determination of pyranose

Claim 5; Page 17-18; 23pp; Japanese.

This invention describes a novel reagent containing pyranose oxidase for the determination of pyranose which can be prepared from a fraction precipitated from an aqueous extract of matsutake mushroom by ammonium sulfate precipitation and has an antibacterial activity against at least

CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.  
 CC 50 kD by gel filtration and shows the presence of components of ca.  
 CC 15 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial  
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral  
 CC aqueous solution and in which the above antibacterial activity is  
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral  
 CC aqueous solution. The invention also describes a method for the  
 CC determination of pyranose in a sample in which the above reagent for the  
 CC determination of pyranose is reacted with pyranose in the sample and the  
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a  
 CC method for diagnosing a disease accompanied by abnormality in sugar  
 CC metabolism in which the above reagent for the determination of pyranose  
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed  
 CC is reacted with an enzyme to develop a color, and a kit for pyranose  
 CC analysis or the diagnosis of a disease accompanied by abnormality in  
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent  
 CC for diabetes mellitus. This sequence represents the Tricholoma matsutake  
 CC pyranose oxidase protein which is described in the method of the  
 CC invention.  
 XX  
 SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;  
 Best Local Similarity 40.1%; Pred. No. 7.7e-89;  
 Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGOIPKNAIHEHYGNDGV-----DVFIAGSGPIGATYAKLCVE--AGLRVVMVEIGAAD 70  
 DB 9 kindllqrsqgdltsqdeivhytdvfiagspiactyarhiidntsttkvymaelsqsd 68

QY 71 SFYAVNAEEGTAVPVPGYHKHNEIEFQKIDRFNVNKGALQOQVSPVRNQNVTLPDG 130  
 DB 69 n-----pvi-gahhrnsikfkdtdkfvniingalqpsidpsdyqptlava 115

QY 131 AWSAPPGSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHWTCTPRTHPPMESL 186  
 DB 116 aw-appldpaegqlvimghnnpqeaaglnlpgsavtrtvggmawhwtacphtphd----- 168

QY 187 GIGRKLNDPAEDDKENWELYSERIEIGTSTKEFDSIRHTLVLSLDAYKDRQRF 246  
 DB 169 -----eervnnpv-dkqefdaaleraktlnvhsdyddsiqivvketlqtl-dasrgv 222

QY 247 RPLPLACHRLKNAPYEVHWSAENLFHSIYNDKQKFLTLTNHRCRLALTGYERKI 306  
 DB 223 tllplgveritdnplyvtgtad----tvlgdvkpsprfvltetrvtckfivsetnptqv 278

QY 307 GAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDSLPN 366  
 DB 279 vaallrnl-----ntsn--delvvaqsfviacgavctpqilwns-----nirph-----a 321

QY 367 LGRIYTEQPMAPFCQIVLRQEFVSDVDDPYGLPMWKEAQAHIKPNPTDALPFPDPDEP 426  
 DB 322 lgrlysegsmtfcqlvlrslvdsiatdpr-----faakeahhkhnpdvlpipthepep 377

QY 427 QVTPFTEHPHWHQIHRDAFSYAGVGEVDSRVIVDLRWFGATDPEANNLIVF----- 480  
 DB 378 qvmyptsdtsfwhvqvrh--yafgvgpkadprvvvdrlrfgksdiveenrvtfgnpkl 435

QY 481 ---QNDVQDGYSMPOPTTRYPSTASNVARKKMDMCEVSNLGGYLPSTPPQPMDEL 537  
 DB 436 rdeagvtdtygmpqptfhvkrtnadgdrdrmmndmtnvanilggylpgsypqfmapgl 495

QY 538 ALHLGAGTTRIGFDKATTVADNNLSVDMFANLYVAGNGTIRTGFGFNGPNTLTSCHAIK 597  
 DB 496 aqhiqgttrigtddqtsvadpctskvnmfdnlwvgngcipdatacnptrtsvayaikgae 555

QY 598 SIINFL 603  
 DB 556 avvsyl 561

RESULT 8

AAV81952  
 ID AAV81952 standard; Protein; 564 AA.  
 XX  
 AC AAV81952;  
 XX  
 DT 07-JUL-2000 (first entry)  
 XX  
 DE Trichoderma derived antifungal protein sequence.  
 XX  
 KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;  
 KW growth inhibitor; plant pathogenic fungi; antibacterial agent;  
 KW N-terminal fragment.  
 XX  
 OS Trichoderma matsutake.  
 PN WO200014242-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 19-AUG-1999; 99WO-JP04441.  
 XX  
 PR 08-SEP-1998; 98JP-0270606.  
 XX  
 PA (NTSB ) JAPAN TOBACCO INC.  
 PA (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.  
 XX  
 PI Takakura Y, Kuwata S, Ohta S;  
 XX  
 DR WPI; 2000-256990/22.  
 DR N-PSDB; AAA07403.  
 XX  
 PT Mushroom-derived antibacterial protein against plant pathogenic fungi  
 PT of rice, with activity and thermal stability, obtainable cheaply on  
 PT large scale, useful in agriculture  
 XX  
 PS Claim 3; Page 41-45; 52pp; Japanese.  
 XX  
 CC This sequence is the Trichoderma matsutake antibacterial protein of  
 CC the invention. The protein has activity against at least Pyricularia  
 CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an  
 CC aqueous extract of a mushroom precipitated by the ammonium sulphate  
 CC precipitation method. The protein has a molecular weight of about 210 kD  
 CC as determined by the gel filtration method, includes components of about  
 CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous  
 CC neutral solution at 60 degrees C for 10 minutes but with loss of  
 CC antibacterial activity after heating in the solution at 80 degrees C for  
 CC 10 minutes. The protein is used for inhibiting the growth of plant  
 CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.  
 CC It is useful in treating rice plants, and is applicable in agriculture as  
 CC an antibacterial agent. The protein has activity at relatively low  
 CC concentrations, and can be produced at low cost on large scale.  
 XX  
 SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;  
 Best Local Similarity 40.1%; Pred. No. 7.7e-89;  
 Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGOIPKNAIHEHYGNDGV-----DVFIAGSGPIGATYAKLCVE--AGLRVVMVEIGAAD 70  
 DB 9 kindllqrsqgdltsqdeivhytdvfiagspiactyarhiidntsttkvymaelsqsd 68

QY 71 SFYAVNAEEGTAVPVPGYHKHNEIEFQKIDRFNVNKGALQOQVSPVRNQNVTLPDG 130  
 DB 69 n-----pvi-gahhrnsikfkdtdkfvniingalqpsidpsdyqptlava 115

QY 131 AWSAPPGSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHWTCTPRTHPPMESL 186  
 DB 116 aw-appldpaegqlvimghnnpqeaaglnlpgsavtrtvggmawhwtacphtphd----- 168

QY 187 GIGRKLNDPAEDDKENWELYSERIEIGTSTKEFDSIRHTLVLSLDAYKDRQRF 246  
 DB 169 -----eervnnpv-dkqefdaaleraktlnvhsdyddsiqivvketlqtl-dasrgv 222

QY 247 RPLPLACHRLKNAPYEVHWSAENLFHSIYNDKQKFLTLTNHRCRLALTGYERKI 306  
 DB 223 tllplgveritdnplyvtgtad----tvlgdvkpsprfvltetrvtckfivsetnptqv 278

QY 307 GAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDSLPN 366  
 DB 279 vaallrnl-----ntsn--delvvaqsfviacgavctpqilwns-----nirph-----a 321

QY 367 LGRIYTEQPMAPFCQIVLRQEFVSDVDDPYGLPMWKEAQAHIKPNPTDALPFPDPDEP 426  
 DB 322 lgrlysegsmtfcqlvlrslvdsiatdpr-----faakeahhkhnpdvlpipthepep 377

QY 427 QVTPFTEHPHWHQIHRDAFSYAGVGEVDSRVIVDLRWFGATDPEANNLIVF----- 480  
 DB 378 qvmyptsdtsfwhvqvrh--yafgvgpkadprvvvdrlrfgksdiveenrvtfgnpkl 435

QY 481 ---QNDVQDGYSMPOPTTRYPSTASNVARKKMDMCEVSNLGGYLPSTPPQPMDEL 537  
 DB 436 rdeagvtdtygmpqptfhvkrtnadgdrdrmmndmtnvanilggylpgsypqfmapgl 495

QY 538 ALHLGAGTTRIGFDKATTVADNNLSVDMFANLYVAGNGTIRTGFGFNGPNTLTSCHAIK 597  
 DB 496 aqhiqgttrigtddqtsvadpctskvnmfdnlwvgngcipdatacnptrtsvayaikgae 555

QY 598 SIINFL 603  
 DB 556 avvsyl 561



PF 13-FEB-1998; 98JP-0031452.

PR 13-FEB-1997; 97JP-0029275.

XX (MOMO-) MOMOYA KK.

PA (NORQ) NORINSUISANSHO SHOKUHIN SOGO.

XX WPI; 1999-074153/07.

XX An anti-tumour protein - prepared by culture of host cell  
PT transformed by vector containing base coding sequence

XX Claim 5; Page 12; 15pp; Japanese.

XX The present sequence represents a peptide derived from an antitumour  
CC protein. Nucleic acid encoding the antitumour protein is isolated  
CC from Tricholoma matsutake.

XX SQ Sequence 51 AA;

Query Match 5.8%; Score 192; DB 20; Length 51;

Best Local Similarity 62.7%; Pred. No. 1.6e-10;

Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 405 VAQHIKNTDLPFPRDPEQVTPTEHPWHTQIHRDAFSYCAVGPE 455

Db 1 veahkkkdpdvlplpfhepeqvmipytstsfpwvqhvrdafsygdvgpk 51

RESULT 11

AAY05719

ID AAY05719 standard; Protein; 615 AA.

XX AC AAY05719;

XX DT 19-JUL-1999 (first entry)

XX DE Gluconate dehydrogenase subunit I.

XX KW Gluconate dehydrogenase; membrane-bound; 2-keto-gluconate;  
KW vitamin C; ascorbic acid.

XX OS Erwinia cyripedii.

XX PH Key Location/Qualifiers

FT Peptide 1..22 /note="signal peptide"

FT Protein 23..615

FT Binding-site 30..60 /note="mature protein"

FT /note="FAD binding motif"

XX WO9915673-A1.

XX PD 01-APR-1999.

XX PF 25-SEP-1998; 98WO-KR00296.

XX PR 25-SEP-1997; 97KR-0048802.

XX (KOAD) KOREA ADV INST SCI &amp; TECHNOLOGY.

XX (KOMI-) KOREA MICROBIAL TECHNOLOGY INC.

XX PI Pan JG, Yum DY;

XX WPI; 1999-254718/21.

XX DR N-PSDB; AAX25345.

XX New isolated gluconate dehydrogenase

XX Claim 13; Page 22-23; 42pp; English.

XX

CC The present sequence represents the 65 kDa subunit I (dehydrogenase  
 CC subunit) of a novel membrane-bound gluconate dehydrogenase (GADH)  
 CC of Erwinia cyripedii ATCC 29267. This GADH is composed of 3  
 CC subunits, the dehydrogenase, a cytochrome c (see also AAY05720) and a  
 CC third subunit (see AAY05721). The dehydrogenase subunit includes an  
 CC FAD-binding motif, suggesting that GADH is a flavoprotein. The gene  
 CC (see AAX25345) encoding the GADH subunits has been isolated. The GADH  
 CC can be used for the production of 2-keto-D-gluconate from glucose or  
 CC D-gluconate in high yields and free of ketogluconate metabolism. A  
 CC claimed method for production of 2-keto-D-gluconate involves  
 CC expression of the GADH gene in a microbial host. Novel recombinant  
 CC E. coli JM109 (pGA313) KCTC 0521BP is claimed. Degenerate DNA  
 CC oligonucleotide probes derived from amino acid sequences of GADH,  
 CC subunit I, II and III are also claimed.

XX SQ Sequence 615 AA;

Query Match 5.8%; Score 191.5; DB 20; Length 615;

Best Local Similarity 21.1%; Pred. No. 1.3e-08;

Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDVFTAGSGPGATYAKLCVEAGLRVWVEIGA-----ADSFYAVNAEGTAVPVPGVH 90

Db 29 vdavvvgfwagalmakelteaglnvalergphrdtydpdgaypqsldelt-----yn 81

QY 91 KKNETEFOKIDIDRFVNVIKGAQQVSVPRNQNVTLDPGAWSPGSSAISNGKNPHQR 150

Db 82 irkki-fq-diskstvtirhdasqtavpyr----- 109

QY 151 EFENLSAEAVTRGVGGMSTHTCTSPRIHPMESLPG-----IGRPKLSNDPAEDD--KE 203

Db 110 ---qlaaflptgtgtgaglhwsghvfrvdpvelnhrshyeargknfipegtlqdfgvs 166

QY 204 WNEL---YSEARLIGTSTKEPDESIRHTLVLSQDAY--KDRQIRPRLP-----IAC 253

Db 167 ynelepfddqackvftgtsaw--tikgmigkekggnfyapdrsd-f-plpaqrtysa 223

QY 254 HRLNAPEVVEWH-----SAENLFHSIYNDD----- 279

Db 224 qlfaqaasvgyhydpmsantsgpytntcyagmgpcnfcgycsyacymyskaspvni 283

QY 280 ----KQKLFLLTNHRTRLALTGGEKKIGA-----AEVRNLLATRNPSQ--LDSYIM 329

Db 284 lpalrqepkfelrnnayvlrvnltdgkkratgvtlydggrevv---qpadvilsafqf 340

QY 330 AKVYVLASGATGNP-QILYNSGFSGLQVTPRNDSLIPNL-GRYTEQPM---AFQCIIVLR 384

Db 341 hnvhlmlisgigqypnptnegvvgrnfaygnistlkalfdknttnpfifagggagvavd 400

QY 385 QEFVDSVRDDPYGL---PWWKEAVAQHIKNTDLPFPRDPEPQVTPTEHPWHT 440

Db 401 dfnadnfdhgygfvvggspfw-----vnaqgtkpvsgltpkgtpn-----wgs 444

QY 441 QIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPEANNLLVFDQVDQGYGSM--OPTFFRYR 498

Db 445 qwk-----aavadtynhhismdah--gahqsyranylldpnyknyvgqpllrmtdwq 496

QY 499 PSTASNVR-ARKMMADCEV--ASN-----LGGYLTSPFPQFMDPGL--ALHLAGTRTIGF 549

Db 497 d-----ndirmagfmvgmrkitekampkmlig--akgpgthfdttvygtthmssgalnge 551

QY 550 DKATTVADNNSLVWDFANLYVAGNGTIRTGENTPT 585

Db 552 dpktsavvrylqswdvpnvfvpgasafpgglgynpt 587

RESULT 12

AAB10459

ID AAB10459 standard; Protein; 67 AA.

XX AC AAB10459;

XX



```
XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX XX
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISB ) JAPAN TOBACCO INC.
XX PA (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX PI Takakura Y, Kuwata S, Inoue Y;
XX DR WPI; 2001-281598/29.
XX XX
XX PT Antibacterial protein and encoded gene isolated from Lyophyllum
PT shimeji, with activity against plant pathogenic bacteria, applicable in
PT agriculture e.g. rice cultivation at low concentration, produced at low
PT cost on large scale.
XX PS Example 2; Page 24; 52pp; Japanese.
XX CC The present sequence is part of an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia orizae and Rhizoctonia
CC solani at a relatively low concentration. P. orizae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
XX CC The antibacterial protein can be produced at low cost on a large scale.
XX SQ Sequence 34 AA;
XX
XX Query Match 5.1%; Score 168; DB 22; Length 34;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-08;
XX Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 211 AERLIGSTKFEDESIRHTLVLSLQDAYKDRQR 244
XX Db 1 aerlgtstkefdesirhtlvrlslqdaykdrqr 34
XX
XX RESULT 15
XX AAB97031
XX ID AAB97031 standard; Peptide; 30 AA.
XX AC AAB97031;
XX DT 20-JUL-2001 (first entry)
XX DE Lyophyllum shimeji antibacterial protein fragment #1.
XX KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX KW Pyricularia orizae; Rhizoctonia solani; rice pathogen.
XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISB ) JAPAN TOBACCO INC.
XX PA (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX XX
```

```
PI Takakura Y, Kuwata S, Inoue Y;
XX DR WPI; 2001-281598/29.
XX DR N-PSDB; AAF99980.
XX XX
XX PT Antibacterial protein and encoded gene isolated from Lyophyllum
PT shimeji, with activity against plant pathogenic bacteria, applicable in
PT agriculture e.g. rice cultivation at low concentration, produced at low
PT cost on large scale.
XX PS Claim 2; Page 32; 52pp; Japanese.
XX CC The present sequence is part of an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia orizae and Rhizoctonia
CC solani at a relatively low concentration. P. orizae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
XX CC The antibacterial protein can be produced at low cost on a large scale.
XX SQ Sequence 30 AA;
```

```
Query Match 4.9%; Score 162; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 76 NAEETAVPYVPGYHKNEIEFQKDIRFV 105
Db 1 naeegtavpyvpgykhkneiefqkdidrfv 30
```

```
Search completed: September 3, 2002, 15:56:34
Job time: 3533 sec
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	1142.5	39.4	623	3	P79076	P79076 coriolus ve
2	170.5	5.9	573	16	Q9P190	Q9P190 campylobact
3	170.5	5.9	551	2	Q9RH54	Q9RH54 pantoea agg
4	166.5	5.7	579	16	Q9A776	Q9A776 caulobacter
5	159	5.5	615	2	O34214	O34214 pectobacter
6	158.5	5.5	529	17	Q9HQ88	Q9HQ88 halobacter
7	157	5.4	545	2	Q9LBA8	Q9LBA8 gluconobact
8	148	5.1	553	2	Q9XCRO	Q9XCRO pantoea cit
9	144	5.0	722	16	Q9RZ76	Q9RZ76 deinococcus
10	141	4.9	523	16	Q98C76	Q98C76 rhizobium l
11	141	4.9	591	16	Q911K8	Q911K8 pseudomonas
12	128.5	4.4	748	10	Q9M0H4	Q9M0H4 arabidopsis
13	128.5	4.4	748	10	Q94BP3	Q94BP3 arabidopsis
14	127	4.4	499	16	Q988P2	Q988P2 rhizobium l
15	117.5	4.1	620	3	Q12623	Q12623 humicola gr
16	116	4.0	502	5	O18672	O18672 caenorhabdi

Db 198 LVTAKESYKTKTDGDFKESIRNNLVNKLAEYKG-QRDFQIPIAATR--RSPTFEVWS 254  
QY 192 SAENLF--HSIYNDKQKKLFTLLNHRCTRRLALTGGYEKKIGAAEVRNLLATRNPSOL 249  
Db 255 SANTVEDLQNRNTDAPNERFNLPVACERV-----VRN-----TSNSEI 295  
QY 250 DSY-----IMAKVYVLAGAIGNPOLYNSGSL-QVTPRN-DSLIPNLGRY 295  
Db 296 ESLIHIDLSIGRFEIKADVFVLTAGAVHNAQLLVNSFGQLGRPPANPPQLPSLGSY 355  
QY 296 ITEOPMAFCQIVLROEFVDSRDDPY--GLP-----WKKEVAQOH 333  
Db 356 ITEQSLLVFCQIVLROEFVDSRDDPY--GLP-----WKKEVAQOH 333  
QY 334 IAKNPTDALPIPRDEPOVTTFFTEHPWHTQIHRDAFSYGAVGEVDSRVIVDLRWFG 393  
Db 416 MQHQEDPLPIPEDEPOVTTFFTEHPWHTQIHRDAFSYGAVGEVDSRVIVDLRWFG 393  
QY 394 AYDPEANLLVFNQVQDGYSMQPTFRYR-PSTASNVRAKMMADMCEVASNLGGYLP 452  
Db 476 RTEPKENKLFSDKITDTYNNPQPTDFRFPAGRTSKEAEDMTDMCVSAKIGGFLPG 535  
QY 453 SPOFMDPGLALHLAGTTRIGDKA--TTVADNNSLVDFANLYVAGNCTITGTGENTPT 510  
Db 536 SLQPMPEGLVHLGGTHMGDFEQEDKCCVNTDSRVFGFKNLFGLGGCGNIPYAGNPT 595  
QY 511 LTSMCHAKSARSIIIN 526  
Db 596 LTAMSLAISKCEVIKN 611

RESULT 2  
Q9PI90 PRELIMINARY; PRT; 573 AA.  
AC Q9PI90;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PUTATIVE OXIDOREDUCTASE SUBUNIT.  
GN CJ0415.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,  
RA Bagham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
RA Jagals K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences."  
RL Nature 403:665-668(2000).  
DR EMBL; AL139075; CAB74251.1; -  
KW Complete proteome.  
SQ SEQUENCE 573 AA; 63686 MW; 7BEB522A779A5F94 CRC64;

Query Match 5.9%; Score 170.5; DB 16; Length 573;  
Best Local Similarity 20.9%; Pred. No. 7.7e-06;  
Matches 111; Conservative 61; Mismatches 237; Indels 123; Gaps 21;  
QY 77 FENLSAEAVTRGVGGMSTHTWCSTPRIPHPMESLPGICRP-----KLSNDPABDD-----K 127  
Db 84 YKMGSLFLGNNVGGAGVHWGWTFRFMPYDFEIQTLKQRYGNKGLNDYTLQDMGVYK 143  
QY 128 ENNELYSEARLIGTSTKEF-----DESIRHTLVLSLQDAYKDRQIFR 172  
Db 144 DMEPYDFEKTGCGSGEPNLAEMKAGFRSPYQPEPLENTKMLKRFESAASSNLHTY 203

QY 173 PLPLA-----CHRLKNAPEYVEHSAENLFHSIYNDKQKKLF 210  
Db 204 RLPASNSKGGYTPDQDLAPCQYCAICERFG-----CEYGAKASPLNTVIFKAMSTGY 258  
QY 211 TLLTHNRCTRRLALTGGYEKKIGAAEVRNLLATRNPSOLDSDYIM-AKYVVLASGAIGNDQ 269  
Db 259 TIRTYSNVYQIL-----KDGKVTGVKFDVTRT-----MKEYIQADIVVLTSYFNNAK 308  
QY 270 ILYNSGFSGLQVTP-----RNDSLIPNLG--RYITEOPMAFCQIVLROEFVDSRDD 319  
Db 309 LLWVNI-GEQDPTKGTGLGRNYCYQNMGTATFAFDEQWTF-----MGSGALGTTSD 363  
QY 320 PYG--LPWKE-----AVAQHIAKNPTDALPIPRDEPOVTTFFTEHPWHTQIHRD 370  
Db 364 FNGDNFHSKREKFLHGMATYSVOLGTRPIQSAPLPAGAP-----TWGAEF-KK 410  
QY 371 AFSYG-----AVGPEVDSRVIVDLRWFGATDEANLLVFNQVQDGYSMQPTFRYRPS 425  
Db 411 ALNYNFTRAITVGGQ-----GASLPHKNNYLSLDPYKDAFGMPLRLTYN-F 457  
QY 426 TASNVRARKMMAD-MCEVASNLGGYLPSPQFMD-----PGLALHLAGTTRIGDKATT 479  
Db 458 TDQDRLAHKFIIDKTAEVAKRMQGVKSKKAYLKDYSVVPYQSTHNTGGTTMGADRETS 517  
QY 480 VADNNSLVDFANLYVAGNCTITGTGENTPTLTSCHAIKASARSIIINTLKGG 531  
Db 518 VVNTYQLQHWADNLPVVGAGNFHNSGYNPTDTVGCALAYRCAEGILKYHKS 569  
RESULT 3  
Q9RH54 PRELIMINARY; PRT; 551 AA.  
AC Q9RH54;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE DEHYDROGENASE SUBUNIT I.  
OS Pantoea agglomerans.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Pantoea.  
OX NCBI\_TaxID=549;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chang K.-I., Yum D.-Y., Pan J.-G., Shin Y.-C.;  
RT "Cloning and expression of a gene cluster encoding three subunits of  
RT membrane-bound 2-keto-D-gluconate dehydrogenase from Erwinia herbicola  
RT ATCC08111 in Escherichia coli.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF068066; AAF21261.1; -  
DR InterPro; IPR000205; NAD\_binding.  
SQ SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAFBF CRC64;

Query Match 5.9%; Score 170; DB 2; Length 551;  
Best Local Similarity 20.3%; Pred. No. 8e-06;  
Matches 109; Conservative 65; Mismatches 209; Indels 154; Gaps 21;

QY 68 NGKNPHREFENLSAEAVTRGVGGMSTHTWCSTPRIPHP-----MESLPGIGRKLSDPAE 124  
Db 93 NVTGNADSFQ-----QGYLRTVGGTTHWAASCHRRHPSDFVMQSKYGVGR----- 139  
QY 125 DDKW-----NEL-----YSEARLIGTSTKEFDESIRHTLVLSLQDAYKDRQIFRPLA 177  
Db 140 ---DWPIGYDELEPYCKAENEIGVAGP--NDPARQSTERS-----QPYFW- 181  
QY 178 CHRLKNAPEYVEHSAENLFHSIYNDKQKKL-----FTLLNHRCTRRL 223  
Db 182 -----DMVPFAHGDNFYASVNVPHGYNLVPYQGRSTRPWEGRPTCCGNNQCIP 233  
QY 224 TGGYEKKIGAAEVRNLLATRNPSOLDSDYIM-----AK 256  
Db 234 IGAMTYNGIHVHE-----RAERNGAVVLAEAIVYKMDTDSNNRITAVHWLDTSGASHKATAK 289



```
QY 257 VYVLASGAINPQILYNSGSLQVTRNDSLIPN-----LGRYTEQPMACQIVLRFQEF 312
Db 290 AFALACNGIETPRLL-----MAANDANPNNGIANSDMVRNMMDHSGFHCSELTKE-- 341
QY 313 VDSVRDDPYGLPMWKEAVAQIAKNPT-DALPIPRDPEQVTPPTTEHPWHQI----- 367
Db 342 -----PWV-----LGKPAQSSCWGYRGD-----FRDYSANKVILNNI 377
QY 368 -----HRDAFSYGAVGPEVDSRV-----IVDLRWFEGATDPEANNLLVFQNDVQDGYSM 415
Db 378 SRVVTATQAAQKGLVGCALDEEIRYRAVHSVDSLSISLEPLDPDENRLTSLKTRKDPHGL 437
QY 416 PQTFRYRPSTASVRARKMADCEVASNLGGYLTPSPQFMDPGLAL--HLAGTRIG 473
Db 438 PCPDIIYDVGDRKGAESAQAQLEHI-----COLFAKEFTISQGLNANNHNGGYIMG 492
QY 474 FDKATTVADNNSLVDFANLYVAGNGTIRTFGFGENPTLTSMCHAISARSISINTLKG 530
Db 493 KNAKEAVVDCNCRADFHENLWLPGGGAIPASVVSNSLTMAALGLKAAHDISLRMKG 549

RESULT 4
ID Q9A7T6 PRELIMINARY; PRT; 579 AA.
AC Q9A7T6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE OXIDOREDUCTASE, GMC FAMILY.
GN CC1634.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT *Complete genome sequence of Caulobacter crescentus.*;
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005839; AAK23612.1;
DR TIGR; CC1634;
DR Interpro; IPR000205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 579 AA; 64397 MW; 7125C4DPAD618F10 CRC64;
```

```
QY 292 -----LGRYTEQPMACQIVLRFQEFVDSV-----RDDPYGLPMWKEAVAQIAKNPTDALP 343
Db 336 ASDQLGRNVMDHHLGAGATGOAPEFADMYFSGRRPNGIYVPRFRN-LGDAASKRSDYLRG 394
QY 344 IPRDPEQVTPPTTEHPWHQIHRDAFSYGAV-----GPEVDSRVIVDLRWFAT 395
Db 395 FGQGGAGGAT-----WERDRGQGGRGFGAARKAALSQPGPW-----MGLSGFGEM 441
QY 396 DPEANNLLVFQNDVQDGYSMQPTFRYRPSTASVRARKMADCEVASNLGGYLTPSP 455
Db 442 LPYADNRVTLNRDVEDKFGLEPLTLMNVTMRDNEAMRMDQAAAAEMLEAAGFONRAHD 501
QY 456 QFMDPGLALHLAGTRIGFRKATTVADNNSLVDFANLYVAGNGTIRTFGFGENPTLTSMC 515
Db 502 NGFAPGLGIHEMGTARMGRDPKTSVLNAHQVHECKNVYVTDGAAMASASCVNPSLTVMA 561
QY 516 HAKSARSISINTLKG 531
Db 562 LTAADAAHVAVRARRKG 577

RESULT 5
ID O34214 PRELIMINARY; PRT; 615 AA.
AC O34214;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE DEHYDROGENASE SUBUNIT PRECURSOR.
OS Pectobacterium cyripedii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=55209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC29267;
RX MEDLINE=98012950; PubMed=9352901;
RA Yum D.-Y., Lee Y.-P., Pan J.-G.;
RT *Cloning and expression of a gene cluster encoding three subunits of
RT membrane-bound gluconate dehydrogenase from Erwinia cyripedii ATCC
RT 29267 in Escherichia coli.*;
RL J. Bacteriol. 179:6566-6572(1997).
DR EMBL; U97665; AAC45885.1;
KW SIGNAL.
FT CHAIN 1 22 DEHYDROGENASE SUBUNIT.
FT CHAIN 23 615
SQ SEQUENCE 615 AA; 67241 MW; B9E1A84FD035609A CRC64;

Query Match 5.5%; Score 159; DB 2; Length 615;
Best Local Similarity 19.9%; Pred. No. 8.3e-05;
Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;

QY 4 EGRAVYVYVGYHKKNEIEFQKIDRFVNI-----KGALQVSVYVRNQVNTLDPGAMS 58
Db 5 ERVSVP-VSGYSRGEGVTYANELKKVDVAVVYVFCWAGAINAKELTAGLNVALERG--- 60
QY 59 APCSSAISNCKNP-----HQREFENLSAEVT-----R 87
Db 61 --PHRDYTPDGAPQSIDELTYNIRKKLFQDLKSTVTIRHDASQTAVPYRQLAALPFGT 118
QY 88 GVGGMSTHWCTSTPRIHPMSELPG-----IGRPKLSNDPAEDD--KEMNEL---YSEAE 137
Db 119 GTGAGLHNSGVHVRVDPVNLNLSHYEARVYKGFIEGFTIQDFGYSYNELEFFFOAE 178
QY 138 RLIGTSTKEFDESIRHTLVLSRLQDAY--KDRQIFRPLP-----LACHRLKNAPEYV 190
Db 179 KVFTSGSAW--TIKGMWIGKEKGNFYADRSSDF-PLPAQKRTYSQAQLFAQAESVGY 235
QY 191 H-----SANELPHSIYND-----KOKKLTFL 212
Db 236 HPYDMP SANTSGPYTNTYGAQMGPFCNFCGCGYACYNYSKASPNVNLPALRQEPKFE 295
```

Qy	213	LTNHRCTRALTGGYEKKIGA-----AEVRNLLATRNPPSQ--LDYSIMAKVYVVLASGAIG	266
Db	296	RNNAYVLRVNLGTGDKKRACTGVYLDGGREW---QPADVLISAFQFHNHVLMLLSGIG	352
Qy	267	NP-QILYNSGFSGLQVTPRNDSLPNL-GRYITEQPM---AFCQIVLRFQFVDSVRDDPY	321
Db	353	QPNYPITNEGVCGRNFAYQNISTLKALFDKNTTNPFIGAGGAGVAVDDFNADFDHGYP	412
Qy	322	GL-----PWKKEAVAQHIAKNPTDALPIPFDPPEQVTPFTTEHPWHQIHRDAFSGAV	377
Db	413	GFVGGSPFW---VNOAGTKPVSCLPTPKGTN-----WGSQWK-----AAV	450
Qy	378	GPEVDSRVIVDLRFWGFATDPEANLLVFQNDQDGYSP--OPTFRRPSTASNVR-ARK	434
Db	451	ADTYNHHISMDAH--GAHOSYRANYLDLPNKVYQGPLLRRTEDWDQ---NDIRMAQF	505
Qy	435	MMADMCEV--ASN---LGSYLPTSPQPMDFCL--ALHLAGTRIGFDPKATTVADNNSL	486
Db	506	MVGKMKKITAMPKMLIG--AKGCTHFTVTYQVTTTHMSGGAIMGEDPKTSAVNRYLQ	563
Qy	487	VWDFANLYVAGNGTIRTFGEENPT	510
Db	564	SWDVPNVFVPGASAFQGLGYNPT	587
RESULT 6			
Q9HQ8R	ID	Q9HQ8R	PRELIMINARY; PRT; 529 AA.
AC	Q9HQ8R		
DT	01-MAR-2001	(TEMBLrel. 16, Created)	
DT	01-MAR-2001	(TEMBLrel. 16, Last sequence update)	
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)	
DE	VNG1035C.		
GN	VNG1035C.		
OS	Halobacterium sp. (strain NRC-1).		
OC	Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;		
OX	Halobacterium.		
OC	NCBI_TaxID=64091;		
RN	[1]		
RX	SEQUENCE FROM N.A.		
RP	MEDLINE=20504483; PubMed=11016950;		
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,		
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sirogna J.,		
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,		
RA	Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,		
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,		
RA	Isenbarger T.A., Peck R.F., Ponischroder M., Spudich J.L., Jung K.-H.,		
RA	Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,		
RA	Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;		
RT	"Genome sequence of Halobacterium species NRC-1."		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).		
DR	EMBL; AF005037; RAG19445.1; -		
DR	InterPro; IPR000425; MIP.		
DR	InterPro; IPR000205; NAD_binding.		
DR	PROSITE; PS00221; MIP; UNKNOWN_1.		
SW	Complete proteome.		
Q9	SEQUENCE 529 AA, 56633 MW; FAED92051336225D CRC64;		

```
Query Match          5.5%; Score 158.5; DB 17; Length 529;
Best Local Similarity 22.3%; Pred. No. 7.2e-05;
Matches 117; Conservative 44; Mismatches 194; Indels 169; Gaps 22;
```

Db	136	HSVPN	-----ARLSAGRETRRACVGYTCQPCVPSGAKYDATVHVHVDATDAGARVID	237
Qy	233	AAEVRNL	-----LATRNPSSQLDSYIMAKYVVLASGAIGNPQIL	272
Db	238	EAPVQRLEHAAGDRVTCGAVYATPDGTTHRQS	---ATEFVLAAGGCIETPRLLLLSDSDRY	294
Qy	273	NSGP	---SGLQVTPRNDSLIINLGRYITQEPMAFCQIVLRQFVDSVRDDPYGLPWKEA	329
Db	295	PDGLANSSGL	-----VGRY	311
Qy	330	VAQHLAKNPDTALPIPRPEPQVTPFTEHPHWTQIHRDAFSYGAV	-----GPEV	381
Db	312	---HLFEACAGGTLDPETRQNHVGFNT	---TESHQY	363
Qy	382	DSRVIVLURWFGA	-----TDEANLLVFQNDYQDGYSPMQPTFR	421
Db	364	AEMALSGDDGWDGMCMDRIRDASGTHIAVGGLEVEQCPRENKRVLPHTDTHVGNVPDVV	423	
Qy	422	YRPSTASNVARKKMWADMCENVASNLGGYLP	-TSPQFQMDPGLALHLAGTTTRIGFDKATVV	480
Db	424	WLSLAYERRTIERANEIQREILTELGADIEWTVGPE	-DTGPAFHMTTTRMGTDPAESV	481
Qy	481	ADNNSLVWDFANLYVAGNGTIRTGEGENTPLTSMCHAIKSARSI	524	
Db	482	VDPRLTRHDLNSLVASSSVFPTAGAMNPTLTIAALAKAADHI	525	
RESULT	7			
ID	Q9LBK8	PRELIMINARY;	PRT;	545 AA.
OC	Q9LBK8;			
DT	01-OCR-2000 (TRENBLrel. 15, Created)			
DT	01-OCR-2000 (TRENBLrel. 15, Last sequence update)			
DT	01-OCR-2001 (TRENBLrel. 18, Last annotation update)			
DE	SOLUBITOL DEHYDROGENASE LARGE SUBUNIT.			
GN	SLDL.			
OS	Gluconobacter oxydans (Gluconobacter suboxydans).			
OC	Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;			
OC	Gluconobacter.			
OX	NCBI_TaxID=442;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=IFO 3254;			
RA	Yoshikawa K., Saito Y., Ishii Y., Noguchi Y., Soeda S.;			
ET	PCT Patent Publication No. WO9920763A1.			
RL	Submitted (FEB-2000) to the EMBL/genbank/DBJ databases.			
DR	EMBL; AB039821; BAA92378.1; .			
DR	Interpro; IPR000205; NAD_binding.			
SQ	SEQUENCE 545 AA; 60076 MW; 4C0A4F3C56950E22 CRC64;			

	Query Match	5.4%;	Score 157;	DB 2;	Length 545;
	Best Local Similarity	19.7%;	Pred. No.	0.0001;	
	Matches 116;	Conservative	69;	Mismatches	138;
				Indels	206;
	Gaps				25;
Qy	34	KGALQQVSVVRNQNVPTLDPGMSAPPSSAISNG----	KNPHOREPENLSAAEAVTGV	89	
				: : :	
Db	58	KGAQLPPVPV-----PNAHPDDOGSPNGYLHTTGPDAAYQ----	OGLRVV	102	
		:	:	:	
Qy	90	GGMTHWTCSPTRIHP---PMESLPFGIGRPKLSNDPADDDKEWNELYSSEARSLITGTSTKE	146		
		: : :	: : :	: : :	
Db	103	GTTTHWAGCAWRYPDSPDFELHSRYGVGR----DWAIKYDLDLEFPFYQAEMVMGVAGPN	157		
		: : :	: : :	: : :	
Qy	147	FDESIRHTLVLRSLQDAYKDKORIF--RPLPLACHRLKNAPEYVEWHSANLFHISIYND	204		
		: : :	: : :	: : :	
Db	158	MD-----VDLLGPSRHNPMEVPLS-----YCAD	183		
		: : :	: : :	: : :	
Qy	205	KOKKFLTLTNHR-----CTRLATGG-----YEKKI	231		
		: : :	: : :	: : :	
Db	184	QFRKLIEKTYNRYVHEPQAANTRPYDKRPTCEGNNNCMPCICGAMYNGIHSYNHAEEA	243		
		: : :	: : :	: : :	
Qy	232	GAAEVRLNLATRNPSOLDSYIM-----AKVVVLASGAI-----	265		
		: : :	: : :	: : :	

Db 244 GARIPNAVVRLETDASNKVVPVNYDDPKNSHRVTGFFVVAANCIESAKLLLLSAD 303  
QY 266 -GNPQILYNSG-----FSLQVLT--PRNDSLIPNLGRIYITEQPMFACQIVLRQEFV 313  
Db 304 DKNPRGIANSSDQVGRNMMDHTGVQLSFGMSGNDLSLWPGRGLLT-----SII 350  
QY 314 DSVRDDPYGLPWKE--AVAQHIKAKPTDALTPIPRDEPQVTPPTFEHPWHTQIHRDA 371  
Db 351 DSFRDG-----PWRSERGAYLVHMYD-----DNOVDFAATGL-----A 382  
QY 372 FSYGAVGPEVDSRVID-----LRWF-----GATDPEANLLVFQNDVODGYSMPOPTFRYR 423  
Db 383 IAKGVCKELEEQIRYGSRAVRULFSNNEGTDADPD--NRUTLSKTHKDVIGIPIPEVYIK 440  
QY 424 ---PSTASNVRAKMMADCEVASNLG-----GYLPTSPQFMGPGGLALHLAGTTRIGFD 475  
Db 441 LPEYTVKSCDHTKELFELMALMSGTDPOWTKGYFP-----QCHPSGSGTINGTD 489  
QY 476 KATTVADNNSLVDFANLYVAGNTITRTGCGENPTLTSMCHAISAKSI 524  
Db 490 PTNSVVDGECRTHDHNLFVARSAVSSVGTGNTITLTIGALALRAAASL 538

RESULT 8  
Q9XCR0 ID Q9XCR0 PRELIMINARY; PRT; 553 AA.  
AC Q9XCR0;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE 2-KETO-GLUCONATE DEHYDROGENASE SUBUNIT.  
GN KDBG.  
OS Pantoea citrea.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Pantoea.  
OX NCBI\_TaxID=53336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=10568;  
RX MEDLINE=20200361; PubMed=10735866;  
RA PuJol C.J., Kado C.I.;  
RT "Genetic and Biochemical characterization of the pathway in Pantoea citrea leading to pink disease of pineapple."  
RL J. Bacteriol. 182:2230-2237(2000).  
DR EMBL; AF131202; AAD4706.1; -  
DR InterPro; IPR000205; NAD\_binding.  
SQ SEQUENCE 553 AA; 59939 MW; F2A60424416B8C3F CRC64;

Query Match 5.1%; Score 148; DB 2; Length 553;  
Best Local Similarity 20.6%; Pred. No. 0.00062;  
Matches 114; Conservative 59; Mismatches 222; Indels 160; Gaps 21;  
QY 52 LDPCANAPGSSAISGNKPHQREFNLNLSAEATRVGGMSTHTWTSTPRIHPP---ME 108  
Db 81 LAPAPLYFPNNY--NVTGPSAGSFQ---QGYLRTVGGTTHWAASCRHHPSDFVMK 134  
QY 109 SLPGIGRP-KLSNDPADDKWENLYSEAERLIGT-----STKEFDESIRHTLVLR 158  
Db 135 SKYGVGRDWFISYDEME---PW---YCAEYEGVAGPSDMSQSPSRSPRYPMQWVPF 188  
QY 159 SLQDAY-----KDRQIRFP---LPLAC-----HRLKN 183  
Db 189 AHGDTYFASVVPNHYNLVPIQGRSTRPWEGRPVCCGNNNCQIPICIGAMNGIHIER 248  
QY 184 APEVVEHSAENLPHSYNDOKKLLTLTNHRCR---LALUGGYEKKIGAAEVRNLL 240  
Db 249 AESGAYVLAESVYKIDTD-----NNRVTAHVHLDNOGASHKATG----- 290  
QY 241 ATRNPSSQLDSYIMAKVYVVLASGAIGNPQILYNSGSLQVTPRNDLSLPN---LGR-- 294  
Db 291 -----KAFALACNGIETPRLLQA-----ANKANPTGIANSMDWGRNM 329

QY 295 -----YITEQPMFACQIVLRQEFVDSVRDDPYGLPWWKEVAQHIKAKNPTDALPIP 345  
Db 330 MDHSGFCHCSFTEEPVWLGRGPAQSSCMVGPDRDGAFRSEY---SANKMILNNISRVVPAT 386  
QY 346 FRDEPQVTPPTFEHPWHTQIHRDAFYSYAGVGEVDSRVI-----VDLRFWFGATDEA 399  
Db 387 -----KOALAKGLVGRKALDDEIRYSIHGVDLSISLEPLPDP 423  
QY 400 NNLVLFQNDVODGYSMPOPTFRYRPFSTASNVRAKMMADCEVASNLGGLYLTSPPOFMD 459  
Db 424 ENRUTLSKTRKDPGLACPDHYDVGYYVRKGATAAHEQIQLHIGSLFNG-----KEFN 476  
QY 460 PGLAL-----HLAGTTRIGFDKATTVADNNSLVDFANLYVAGNTITRTGCGENPTLTSMC 515  
Db 477 ITTALNANNHMGITMGSAKDAVDGNCRTFDHENLWLPGGGAIPASVYVNSTLSMAA 536  
QY 516 HAIKSARSIIINTLK 529  
Db 537 LGLKAAHDI SLRMK 550

RESULT 9  
Q9RZ26 ID Q9RZ26 PRELIMINARY; PRT; 722 AA.  
AC Q9RZ26;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE GMC OXIDOREDUCTASE.  
GN DRA0127.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1."  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001862; AAFI2230.1; -  
DR TIGR; DRA0127; -  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR001199; Cyt\_B5.  
DR InterPro; IPR000172; GMC\_oxred.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00732; GMC\_oxred; 1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
DR PROSITE; PS00191; CYTOCHROME\_B5\_1; UNKNOWN\_1.  
DR PROSITE; PS00624; GMC\_OXRED\_2; 1.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 722 AA; 75376 MW; 450DF1CD1B7596F9 CRC64;

Query Match 5.0%; Score 144; DB 16; Length 722;  
Best Local Similarity 21.1%; Pred. No. 0.0021;  
Matches 107; Conservative 68; Mismatches 196; Indels 136; Gaps 26;  
QY 90 GGMSTHTCTPRIPHPMESLPGICRPKLSN-----DPAEDDKWENLYSAERLIG 141  
Db 277 GGSIVNNSV-----PPRDDI-----RQWASEHGLSDVADPGYD-----RHDAVLERM-G 323  
QY 142 TSTKEFDESIRHTLVLRSLQDAYKDRQIRFPRLPLACHRLKNAPEYVHSAENLPHSYI 201

Db 324 VSEQSDHNGPHQ-----RLVEGADKLGTYTVKA-----ALNLSPEH---YDADKAGHAGF 371  
 QY 202 NDD---KOKKLEF-----LITNHRCTRRLALTGGYEKKIGAAAEVRNLLATRNPS 247  
 Db 372 GDOTGAKOGLTNFLKDAFEAGARILVGTQRAQRLVEDG-----NAAQVSATVTGDETR 426  
 QY 248 QLDSTYMAKVYVLAGAIGNPOLYNSGFSGLQVTPRNDSLPNLGRYITEQPMFACQIV 307  
 Db 427 QIT--VRAPQVVVAGCAETPALLRSIGG-----PAAGRYLRHPAGIVAGI 473  
 QY 308 LRQEFVDSVRDDPYG--LPWKEAQAQIAKNPTDALPIPERDPE-----P 351  
 Db 474 -----YGEDQAWNGP-----POSGILKQFADHENGHGFTIEGVQYGP 511  
 QY 352 QVTT---PPT---EEH-PWHTQIHRDAFSYGAVGPEVDSRIYVD-----LRWFGATDP-EA 399  
 Db 512 ALMASGLPWTGCEAHRDLMSKFRHATFVSIQDRGCHGQVTVDDGNAVHTYALTDDLDA 571  
 QY 400 NMLLVFQNDVQGYSM-----POPTFRYRPTASNVARKMMADMCVEAS-----NLGGYLP 451  
 Db 572 RN---FRGVTESIIRLHEAAGAEETVALAPGVAPWRRGDDLEAFIGQVAQVPLGAGQTV 628  
 QY 452 TSPQPQMDPGLHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGCGENPFL 511  
 Db 629 FS-----AHQMSARMGSDQTSVADPDGQLHDVPGVWIGDTSAPFTCSGVNPMV 678  
 QY 512 TSMCHAIKSARSIIINTLKGTDGKNTG 538  
 Db 679 SCMALASRTAEKLLAAMEGADGTGSG 705

## RESULT 10

Q98C76 ID Q98C76 PRELIMINARY; PRT; 523 AA.  
 AC Q98C76;  
 DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)  
 DE MLR5266 PROTEIN.  
 GN MLR5266.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP003006; BAB51745.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 523 AA; 58536 MW; F026EA2A6F9E3BE8 CRC64;

Query Match 4.9%; Score 141; DB 16; Length 523;  
 Best Local Similarity 20.9%; Pred. No. 0.0022;  
 Matches 101; Conservative 60; Mismatches 194; Indels 128; Gaps 20;  
 QY 86 TRGVCGMTHWTCSTPRHP---PMESLPGIRPKLSND---PAEDDKENNELYSAEIRLI 140  
 Db 124 TIGLGGGNCNFAQTPFRHPNDLFTKTYGIG-----NDWPISYDDVE--PFYCDAAEII 176  
 QY 141 GTS-----TKFEDESIRHLLVLSLQDAYKDRQIRFP-----LPLA-----CH 179  
 Db 177 SISGDPDMMMLPRSKPFPQPHRMSTPDKMKAAQPDQHFVMPARARVPTAQTSCCA 236

QY 180 RLK--NAPYEVWHSAENLFHSIYNDKQKLFLLTINHRCTRRLALTGGYEKKIGAAEYR 237  
 Db 237 NLRCWLCPPVDAKFTANGLMHVFEHPD-----VSVCLGAEVRRLDQVGG-----TVR 283  
 QY 238 NLLATRNSSQLDSYIMAKVYVLAGAIGNPOLYNSGFSGLQVTPRNDSLPNLGRYIT 297  
 Db 284 SVTFVHDGK---EYQVSGDLFILGANAIQSAAILRSG----- 318  
 QY 298 EQPMFACQILVROEVDVSRDDPYGLPW---WKEAVAQIAKNPTDALPIPRDPEQV 353  
 Db 319 -----LRDEFVGRGLHESYG--WNFEVYLDGVNDFDGTITTGLNFGLYD----- 361  
 QY 354 TTPFTTEHPHTQIHRDAFSYGAVGPEVDSRIYVDLRFMGATDPEANNL-----LVFQ 406  
 Db 362 -GPHRSEH-----AAALVYFENRWQHGMRAEKGLRLQTLPLVVVTE 401  
 QY 407 N--DVQDGYSMQPTFRYRPTASNVARKMMADMCVEASNLGGYLPSPQPMFDPGL-- 462  
 Db 402 NLLDDENFVTLDEDDNAFVSEKAPSDYAVKGMARADKLPELLAPLPVE--RLFDGRIRP 459  
 QY 463 -ALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGCGENPFLTSMCHAIKSA 521  
 Db 460 TESHVQGLRMTGTADSVIDSNMIIHRLNLYVVGTVSTYTPSCSCANPSELTAALSLRAA 519  
 QY 522 RSI 524  
 Db 520 SRI 522

RESULT 11  
 Q91IK8 ID Q91IK8 PRELIMINARY; PRT; 591 AA.  
 AC Q91IK8;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)  
 DE GLUCONATE DEHYDROGENASE.  
 GN PA2265.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.D., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen";  
 RL Nature 406:959-964(2000).  
 DR EMBL; AE004652; AAG05653.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 591 AA; 64716 MW; 9A7012B67ECEB9DE CRC64;

Query Match 4.9%; Score 141; DB 16; Length 591;  
 Best Local Similarity 19.9%; Pred. No. 0.0027;  
 Matches 120; Conservative 63; Mismatches 235; Indels 184; Gaps 22;  
 QY 45 RNQVPTLDPGANW-----SAPPOSSAISNCKNPHQREFEMLSEAVTR 87  
 Db 38 RGENRDYTPDGAYPNTLDYNTYRGLFQNLKSTVSIHGINDTALPYRLSALFPGD 97  
 QY 88 GVGGSTWTCSTPRHPMESLPG-----IGRPKLSNDPAEDDKENNELYSAE 137  
 Db 98 GVGGAGLHWGSHVFRIMPPEELRLRSHYEERYGKKFIPGMITQDYGVSYEELEPHDFAE 157  
 QY 138 RLIGTSTKEFDESIRHLL-----VLRSLQDAYKDRQIRF----- 172

```

Db 158 KVFETS-----GTAHTVTKQVVKGNPFPAADRSDDPPLPALRQVYS--AQLERKAAEEL 209
QY 173 -----PLPLA-----CHRLKNAPEYWEHSAENLFHSIYNDKQKK 208
Db 210 GLHPYDLPAANASGPWNPXGVOMGPNCFGFCGYACMYKASPNL--NILPALRQTP 267
QY 209 LFTLLNHRCTRRLALGGYEKKIGA-----AEVRNLLATRNPSOLDOSYIMAKYVVLASGA 264
Db 268 LFEELRANCNVLYKNLSDGQATGVTYVDAQGREIVQP-----AKLVIISAFQ 315
QY 265 IGNPOLLYNSGSGSLQVTPRNDSLIPNLGRYITEQPMAFQIIVLRQEFVDSVRDPY--- 321
Db 316 FHNVRLLLSGI-GKPYDPTGTGGV--VGKNFAYQNNM-----TIKAFFDKDVHTNPFVGT 368
QY 322 -----GLPWWKEAVAQIAKNAKNTDALLPIPRDPPEQVTTTF 357
Db 369 GGGGVAVDNADNEDHGPLGFGGSPMW-----VNQAGSKPIGGLAVPPCTPS----- 417
QY 358 TEHPHPT---QTHRAFSYGAVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQDYS 414
Db 418 -----WGSQKQAVKDAYTH-----TVSMDAHGSMNTYRNYLDLDPYKDAYG 461
QY 415 MP--OPTERY-----RPTASNVARKMMADMCEVA-----SNLGGVLPSPPOFMDPGL 462
Db 462 QPLLRTFTDKNEIRMSRVTEHMKRIABAMPKALSVSKNGDHFNTRVYQ----- 515
QY 463 ALHLAGTTRIGFDKATTVADNNSLVDFANLYVAGNGTIRTFGENPTLTSCHAIKASR 522
Db 516 TTHLGGGAINSGDPKTSVLNRYLQSDVHNVFVNGASAFPGQGYNPPTGLVAALAYWSAK 575
QY 523 SI 524
Db 576 AI 577

RESULT 12
Q9M0H4 PRELIMINARY; PRT; 748 AA.
AC Q9M0H4
AT Q9M0H4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN AT4G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161573; CAB81445.1;
DR HSSP; P22637; 3COX.
DR InterPro; IPR000205; NAD_binding.
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81959 MW; 13CD183F5940DAC3 CRC64;

```

```

Query Match 4.4%; Score 128.5; DB 10; Length 748;
Best Local Similarity 19.7%; Pred. No. 0.046;
Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;

QY 76 EENLSAEAVTRGVGGMSTHTCTSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSE 135
Db 301 KFMLLAGSAGV---GGTAVNWSAS---IRTPDHVL-----QEM-----SE 334

```

```

QY 136 AERLICHTSTKEF-----DE-SIRHTLVLRSLQDAYKDRQIRFRPLPLACHRL-----K 182
Db 335 GSKIKFGSQEYQASMDDEVIRIGVTERCVKHGFQN-----QVLRKGCERGLQVESVPR 389
QY 183 NAFB-----YVEHSAENLFHSIYNDKQKKLFTLLNHRCTRRLAL-----TGGYEKK 230
Db 390 NSPEDHYCGLCGCGCRAGAKNGTDQTLVDAVENGAVILTGKAERFVLVDNTSSSSNERK 449
QY 231 -----TGAAEVRNLLATRNPSOLDOSYIMAKYVVLASGAIGNPOLLYNSGSGSLQVTPRN 285
Db 450 KRCVGFPAASVGKGKGGK-----FIIEARVTVSSAGSLTLPMLSSGLKN----- 495
QY 286 DSLIPNLGR-----YITEQPMAFQIIVLRQEFVDSVR--DDPYGLPWWKEAVA 331
Db 496 -----PNIGRNLKHPVLMTGWYFPEKDSFGKMEGGIITSVHMND-----TESGC 544
QY 332 OHTAKNPTDALPIPRDPPEQVTTPTTEHPHPTQTHRAFSYGAVGPEVDSRVIVDLRW 391
Db 545 KALENPLIG-PASYAGLSPWVS-----GPDLLKERMIK----- 576
QY 392 FGATDPEANNLLVFQNDVQDYSM--PQPTFRVPRPSTASNVARKMMADMCEVAS---NL 446
Db 577 YGRY-----AHLFALVRDLGSGEVMENEVYRTTKDRENLRAGLRQALRVSVAAGAVEV 632
QY 447 GGY-----LPTSPPOFMDPGLAL-----HLAGTTRIGFDKAT 478
Db 633 GTVRSDDQMKCEAITKEAMEEFLDEVDVAGVGVTGKEYTTFSAHQMSGRMGVTAAE 692
QY 479 TVADNNSLVDFANLYVAGNGTIRTFGENPTLTSCHAIKASRIINTLKGGT 532
Db 693 GALDENGESWEAREGLFVCDGSILPSAVGVNPMITIQSTAYCISSKIIVDSLQNK 746

RESULT 13
Q94BP3 PRELIMINARY; PRT; 748 AA.
AC Q94BP3
AT Q94BP3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN AT4G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carinci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene AT4G28570 (GI:7269712).";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY039977; AAK64154.1;
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81987 MW; E3EB613F5D56A0D1 CRC64;

```

```

Query Match 4.4%; Score 128.5; DB 10; Length 748;
Best Local Similarity 19.7%; Pred. No. 0.046;
Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;

QY 76 EENLSAEAVTRGVGGMSTHTCTSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSE 135
Db 301 KFMLLAGSAGV---GGTAVNWSAS---IRTPDHVL-----QEM-----SE 334

```

```
QY 136 AERLIGSTKPE-----DE-SIRHTVLRLSLQDAYKDRQIRFRLPLACHRL-----K 182
DB 335 GSKIRFGSQEYQSAMDVEYIRIGVTERCVKHGFQ-----QVLRKGCERLQLQVESVR 389
QY 183 NAPE-----YVEWHSANLFSHLYNDKOKKLLFTLLTNRHCTRLAL-----TGGYEKK 230
DB 390 NSPDHYCGLGCGRAGAKNTDQTLVDAVENGAVILTGKAERFVLVDNTSSNERK 449
QY 231 -----IGAAEVNLLATRNPSQLDSYIMAKVYVLASGAIGNPQILYNSGSLQVTPRN 285
DB 450 KRCVGVFASSVGKIGK-----FIIEARVTVSSAGSLTTPMLSSGLKN-----495
QY 286 DSLINLGR-----YITQPMACFOILVRQEFVSVR--DDPYGLPWNKEAVA 331
DB 496 -----PNIGRNKLHPVLMTGWYFPEKDEFSGKMYEGGIITSVHMND-----TESGC 544
QY 332 OHIAKNPTDALPIPRDEPQVTTFTTEHPHWTQIHRDAFSYGAGVEPDSRVIVDLRW 391
DB 545 KAILLENPLIG-PASYAGLSPWVS-----GPDKERMIR-----576
QY 392 FGATPEANLLVFQNDVDGYSM--POPTFRYRPSTASNVARKMADMCEVAS---NL 446
DB 577 YGRT-----AHLFALYRDLGSGEVMMENEVYTTKKDRNLRLAGLRQALRVSYAAGAVEV 532
QY 447 GGY-----LPTSPPOFMDPLAL-----HLAGTTRIGFDKAT 478
DB 633 GTYRSDGQMKCEAITKEAMEEFLDEVDVAGVGVTGYEWTYTFSAHQMGSCRMGVTAEE 692
QY 479 TVADNLSLWMDANLYVAGNCTIRTFGENPTLTSMCHAIKSARSLIINTLGGT 532
DB 693 GALDENGESWEAEGFLVDCGSLPSAVGNPNMTITQSTAYCISSKIVDSLQNK 746

RESULT 14
Q988P2 PRELIMINARY; PRT; 499 AA.
AC Q988P2;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MLR6655 PROTEIN.
GN MLR6655.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003009; BAB52905.1; -
DR InterPro: IPR000205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 499 AA; 54537 MW; 633924AED9688F51 CRC64;
```

```
Query Match 4.4%; Score 127; DB 16; Length 499;
Best Local Similarity 20.68; Pred. No. 0.033;
Matches 118; Conservative 58; Mismatches 223; Indels 174; Gaps 26;

QY 13 GYHKKNEIFQKIDRFVNIKGLQYVFPVNRQNVPTLDPGAWSPGSSAI-----66
DB 37 GEHLRDTPEARDIAIFQNGFYSSEE-----WLAVDGESFLPGNYYY 79
QY 67 --SNGK-----NPHOREFENLSAEAVTRCVGMSTHWTCTPRIHP---PM 107
```

```
DB 80 VGGNSKFFGAMVRYRQEDFNDRD-----MGRSPGWPISYAELEPWYERA 126
QY 108 ESLPGIGRPKLSNDPAEDKKNELYSERLIGTSTKEFDESIRHTVLRLSLQDAYKDR 167
DB 127 EILFGV-RGDARQDTEPR--NRPY-----RYLPVDPDEPATVQRRLQAGIHFA----175
QY 168 QRIFRPLPLA-----CHRLKNAPEYVWHSANLFSHLYNDKOKKLLFTLLTNRHCTRL 221
DB 176 -----SLPLAIDIDAWLRRAKTG-----WDAPPNT--GAGKIDAEGVPLTKALEHPNATL 223
QY 222 ALTGGEYKKGKGAEEVNRNLLATRNPSQLDSYIMAKVYVLASGAIGNPQILYNSGSLQV 281
DB 224 ITGANVQRVLVTDAAGRRVMAAVFIKDGVELSIGADVAFVAGVQSAALLLRS-----276
QY 262 TPRNDSLTPN-----LGR-YITEQPMACFOI-----VLROE---FVDSVRDDPYG 322
DB 277 ---STSVYPNGLGNSSDQIGRNFNMHNTAMLAIDPFRNRNTAVYQKTLGFNDFYNKDFLG 333
QY 323 -LPWNKEAVALHIAKNPTDA---LPPIPRDEPQVTTFTTEHPHWT-QIHRDAFSYGA 376
DB 334 SFPLGNVOLLGHITGNILKANAPLLP-----RWLAGLVARNICYGWFL 375
QY 377 VG---PEVDSRVIVDLRWFGATPDEANLLVFQNDVDGYSMPOPTFRYRPSTASNVRAR 433
DB 376 TSEDLPNPESRVTI-----RNGRIVMNVNRNMGAHETLIR-----RTR 414
QY 434 KMW--ADMCEVASNLGGYLPSPPOFMDPLGLHLAGTTRIGFDKATVADNLSLWMDFA 491
DB 415 AVMRAGFPVVLTRTFGRKTTT-----HOCGARLGSDPNTSVSPDCRSHDIA 463
QY 492 NLYVAGNCTIRTFGENPTLTSMCHAIKSARSI 524
DB 464 NLYVTDASVLTSAVNPAITIALAKAGRAI 496
```

```
RESULT 15
Q12623 PRELIMINARY; PRT; 620 AA.
AC Q12623;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GLUCOAMYLASE (EC 3.2.1.3).
GN GLA1.
OS Humicola grisea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5527;
RN [1]
RP SEQUENCE FROM N.A.
RA Berka R.M., Rey M.W., Thompson S.A., Gray G.L., Carmona C.L.,
RA Power S.D.;
RT "Molecular cloning, analysis and expression of the glal gene encoding
RT a thermostable, raw starch-digesting glucoamylase from the fungus
RT Humicola grisea var. thermoides."
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: M89475; AAA33386.1; -
DR HSSP: P04064; IGAI
DR InterPro: IPR002044; CBD_4.
DR InterPro: IPR000165; Glyco_hydro_15.
DR Pfam: PF00686; CBD_4; 1.
DR Pfam: PF00723; Glyco_hydro_15; 1.
DR PRINTS: PR00736; GLHYDRASE15.
DR ProDom: PD001568; CBD_4; 1.
DR PROSITE: PS00820; GLUCOAMYLASE; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 620 AA; 66525 MW; 54214FF67E20BE0A CRC64;
```

```
Query Match 4.1%; Score 117.5; DB 3; Length 620;
Best Local Similarity 19.08; Pred. No. 0.3; Mismatches 246; Indels 189; Gaps 30;
Matches 124; Conservative 94;
```



**This Page Blank (uspto)**